Lecture Notes in Computer Science 2717
Edited by G. Goos, J. Hartmanis, and J. van Leeuwen

Springer
Berlin
Heidelberg
New York
Hong Kong
London
Milan
Paris
Tokyo

James C. Gee J.B. Antoine Maintz
Michael W. Vannier (Eds.)

Biomedical Image Registration

Second International Workshop, WBIR 2003
Philadelphia, PA, USA, June 23-24, 2003
Revised Papers

Springer

Volume Editors

James C. Gee
University of Pennsylvania, Department of Radiology
3600 Market Street, Philadelphia, PA 19104-2644, USA
E-mail: gee@rad.upenn.edu

J.B. Antoine Maintz
Utrecht University, Department of Computer Science
P.O. Box 80089, 3508 TB Utrecht, The Netherlands
E-mail: twan@cs.uu.nl

Michael W. Vannier
University of Iowa College of Medicine, Department of Radiology
200 Hawkins Drive 3985 JPP, Iowa City, IA 52242, USA
E-mail: michael-vannier@uiowa.edu

Cataloging-in-Publication Data applied for

A catalog record for this book is available from the Library of Congress

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliographie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>.

CR Subject Classification (1998): I.4, I.5, H.3, J.3

ISSN 0302-9743
ISBN 3-540-20343-5 Springer-Verlag Berlin Heidelberg New York

Springer-Verlag Berlin Heidelberg New York
a member of BertelsmannSpringer Science+Business Media GmbH

http://www.springer.de

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Olgun Computergrafik
Printed on acid-free paper SPIN: 10928912 06/3142 5 4 3 2 1 0

Preface

The 2nd International Workshop on Biomedical Image Registration (WBIR) was held June 23–24, 2003, at the University of Pennsylvania, Philadelphia. Following the success of the first workshop in Bled, Slovenia, this meeting aimed to once again bring together leading researchers in the area of biomedical image registration to present and discuss recent developments in the field.

The theory, implementation and application of image registration in medicine have become major themes in nearly every scientific forum dedicated to image processing and analysis. This intense interest reflects the field's important role in the conduct of a broad and continually growing range of studies. Indeed, the techniques have enabled some of the most exciting contemporary developments in the clinical and research application of medical imaging, including fusion of multimodality data to assist clinical interpretation; change detection in longitudinal studies; brain shift modeling to improve anatomic localization in neurosurgical procedures; cardiac motion quantification; construction of probabilistic atlases of organ structure and function; and large-scale phenotyping in animal models.

WBIR was conceived to provide the burgeoning community of investigators in biomedical image registration an opportunity to share, discuss and stimulate developments in registration research and application at a meeting exclusively devoted to the topic. The format of this year's workshop consisted of invited talks, author presentations and ample opportunities for discussion, the latter including an elegant reception and dinner hosted at the Mütter Museum. A representation of the best work in the field, selected by peer review from full manuscripts, was presented in single-track sessions. The papers, which addressed the full diversity of registration topics, are reproduced in this volume, along with enlightening essays by some of the invited speakers. The special lectures were given by David Hawkes from King's College, London, Ron Kikinis from Brigham and Women's Hospital, and Klaus Klingenbeck-Regn from Siemens Medical Solutions. David's dazzling presentation highlighted several of the aforementioned applications of registration that are generating such tremendous excitement in the field. Klaus envisaged that this level of interest will only intensify as innovations in registration methodology are leveraged to advance the conduct of interventional procedures. Against this backdrop of unfettered optimism, Ron reminded the algorithm developers to keep sight of practical considerations in their work and the ever-present challenge of clinical validation of novel results.

As well as convening the primary researchers in the field, this year's meeting also served as a tribute to Ruzena Bajcsy on the occasion of her 70th birthday. Ruzena's landmark research on elastic registration and digital anatomy atlases is the foundation on which the most important work in this discipline is built. She continues to build upon her long and distinguished career in computer science, in her roles as a scientist and academic leader. In the latter position, Ruzena served as the head of the Directorate for Computer and Information Science

and Engineering at the National Science Foundation, where she spearheaded the Information Technology Research Program, before moving to Berkeley to lead the Center for Information Technology Research in the Interest of Society. Prior to these positions, Ruzena was faculty at the University of Pennsylvania, where she founded the renowned GRASP Laboratory. Throughout her career, her influence has extended beyond the reach of her personal research, continually exemplifying the best in scientific investigation.

A special session led by Jim Gee from the University of Pennsylvania gathered several distinguished colleagues who had participated in some of the milestone events that marked Ruzena's career in medical imaging. Martin Reivich, Director of the Cerebrovascular Research Center at the University of Pennsylvania, discussed his group's development of the radiotracer fluorine-labeled deoxyglucose (FDG) – which has been hailed as the "molecule of the 20th century" for its unparalleled and unique impact on the evolution of the field of nuclear medicine – and their demonstration for the first time of FDG-PET's potential for quantifying regional cerebral function such as blood flow and volume. The need to correlate this information with anatomy led to a collaboration with Ruzena and directly motivated her conception of a digital brain atlas that could be individualized to label the neuroanatomy of any subject. Stanley Baum, Editor of Academic Radiology, described the Institute of Medicine's recognition of this seminal development in its citation honoring Ruzena's induction into the institute. Siegfried Stiehl, Dean of the Department of Informatics at the University of Hamburg, and Alok Gupta, Head of the Computer-Aided Diagnosis and Therapy Group at Siemens Medical Solutions, both reflected on the remarkable experience of having Ruzena as a mentor and being the beneficiaries of the profound affection she has for and empathy she shares with her trainees. Eduardo Glandt, Dean of the School of Engineering and Applied Science at the University of Pennsylvania, concluded the tribute by presenting Ruzena with a certificate from the school "in recognition of her seminal research contributions in the field of medical imaging and for her outstanding mentorship of students and young colleagues."

There was an overwhelming level of interest in WBIR 2003, and, despite adjusting upward the attendance limit, we were still forced to turn down many registration requests from all over the world. We were delighted to welcome participants to Philadelphia and hope that you found the meeting an enjoyable and stimulating experience. For readers unable to attend the workshop, we hope that you will find this a valuable record of the scientific program.

June 2003

James C. Gee
J.B. Antoine Maintz
Michael W. Vannier

Acknowledgements

The 2nd International Workshop on Biomedical Image Registration was made possible by the efforts of many people and generous organizations. First, we are indebted to the members of the Scientific Program Committee for their critical reviews that determined the content of the program. We also extend our gratitude to all authors who submitted papers to the workshop.

We gratefully acknowledge the assistance of the dedicated team of students and fellows at Penn who helped with the organization and conduct of the workshop, particularly Tessa Sundaram, Jeffrey Duda, Zhiyong Xie, Lijun Ding, Hui Zhang, Marcelo Siqueira, Brian Avants, and Jonathan Leung.

Finally, we express our appreciation of the financial support provided by the following organizations:

Siemens Medical Solutions
Siemens Corporate Research
National Library of Medicine
Vice Provost for Research, University of Pennsylvania

Organization

Workshop Chairs

James C. Gee	University of Pennsylvania, USA
J.B. Antoine Maintz	Utrecht University, The Netherlands
Michael W. Vannier	University of Iowa, USA

Program Committee

D. Alexander	University College London, UK
M. Analoui	Pfizer Global Research and Development, USA
N. Ayache	INRIA, France
C. Barillot	University of Rennes, France
M. Braun	University of Technology, Sydney, Australia
E. Bullitt	University of North Carolina, Chapel Hill, USA
G. Christensen	University of Iowa, USA
D. Collins	Montreal Neurological Institute, Canada
C. Davatzikos	University of Pennsylvania, USA
B. Dawant	Vanderbilt University, USA
M. Fitzpatrick	Vanderbilt University, USA
F. Gerritsen	Philips Medical Systems, The Netherlands
A. Goshtasby	Wright State University, USA
A. Gupta	Siemens Medical Solutions, USA
P. Hellier	IRISA, France
D. Hill	King's College London, UK
S. Joshi	University of North Carolina, Chapel Hill, USA
S. Kovačič	University of Ljubljana, Slovenia
A. Krishnan	Siemens Medical Solutions, USA
B. Likar	University of Ljubljana, Slovenia
M. Loew	George Washington University, USA
F. Maes	University of Leuven, Belgium
C. Maurer	Stanford University, USA
C. Meyer	University of Michigan, USA
X. Pennec	INRIA, France
F. Pernuš	University of Ljubljana, Slovenia
S. Pizer	University of North Carolina, Chapel Hill, USA
J. Pluim	University Medical Center, Utrecht, The Netherlands
T. Rohlfing	Stanford University, USA
K. Rohr	International University in Germany, Germany
D. Rueckert	Imperial College London, UK
Y. Sato	Osaka University, Japan

F. Sauer	Siemens Corporate Research, USA
D. Shen	University of Pennsylvania, USA
C. Studholme	University of California, San Francisco, USA
P. Thompson	University of California, Los Angeles, USA
A. Toga	University of California, Los Angeles, USA
D. Vandermeulen	University of Leuven, Belgium
B. Vemuri	University of Florida, USA
M. Viergever	University Medical Center, Utrecht, The Netherlands
S. Warfield	Brigham and Women's Hospital, Harvard University, USA
J. Williams	Siemens Corporate Research, USA
C. Xu	Siemens Corporate Research, USA

Additional Reviewers

S. Periaswamy
J. Stoeckel
M. Wolf
P. Cathier
N. Paragios

Table of Contents

Regularization

Novel

Brain

Piecewise Affine

Information Theoretic

Interventional

Applications

Invited Papers

Posters

New Vector Field Regularization Techniques for Nonrigid Image Registration

Pascal Cathier* and Nicholas Ayache

Epidaure, INRIA-Sophia, France

Abstract. The aim of this paper is to propose new linear isotropic regularization techniques for vector fields. In nonrigid registration, most of the regularization techniques treat independently each component of the deformation. However, real materials often have non-zero Poisson ratio, as a stress in one direction would affect the position of the material in every direction. This is especially true for near-incompressible materials such as brain tissue. Therefore, explicit modeling of this property is expected to produce more accurate registration results. In this paper, we propose a new family of isotropic quadratic energies that possess such cross-effects, and derive filters and splines for vector field regularization. Contrary to usual filters and splines (e.g Gaussian filters or thin plate splines), these ones are vectorial, and have parameters that enable them to control the strength of the interaction between coordinates. We show how they can be used for intensity- or landmark-based registration, and finally show how they mix when combining intensity- and landmark-based registration.

1 Introduction

Roughly speaking, there are two main nonrigid registration methods. The first one is the geometric feature based (or landmark-based) method. Some geometric features, such as points, lines, or surfaces, are extracted beforehand in a couple of images to be registered. Then, a geometric distance D (e.g. the Euclidean distance between closest points) is used to match these two sets. The second one is the intensity-based method. The image couple is registered by optimizing a similarity measure D (e.g. the mutual information) between their intensity.

In both cases, it is not sufficient to rely only on the distance D between intensity or geometric features. Without any additional constraint, the motion of a point would be estimated independently of its neighbors, yielding a very noisy and unrealistic displacement field. We should therefore regularize this estimation problem using our *prior knowledge* on the kind of deformation we are looking at.

One popular way of inferring spatial correlation in the estimation is to use a regularization energy R, which, added to the distance D to be minimized, penalize a transformation proportionally to its irregularity. Under some conditions,

* Now with Siemens Medical Solutions USA, Malvern, PA.

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 1–10, 2003.

the use of such an energy can be equivalent to the use of convolution filters or splines, which are very popular in registration.

Most of the time, however, the regularization energy R measures the smoothness of the vector field by measuring the smoothness of its coordinates independently. As a consequence, convolution filters or splines used to estimate the transformation are scalars, and applied to each coordinates independently. This behavior is very unrealistic for real materials, which have commonly a non-zero Poisson ratio – especially near-incompressible tissues such as brain tissue. Therefore we should find a way to introduce, in the regularization energy, a coupling of the coordinates.

The linear elastic energy is the most common example of an energy offering such a coupling, *via* the Lamé coefficients. However the linear elasticity energy may be unsatisfactory for at least two reasons. First, its associated convolution filter has a very strong discontinuity in its derivative, yielding unsmooth displacement fields. Second, its associated spline does not exist in dimension two or higher – even if some substitutes have been designed [3]. Both problems occur because this energy relies on the first order partial derivative. There is therefore a need of higher-order energies having this coupling property.

In this paper, we first propose in Section 2 a new class of quadratic regularization energy that possess such coupling, and that can be of an order as high as desired. Then we study their associated filter in Section 3. Contrary to conventional scalar filters, these vector filters allow coordinates to interact during deformation. We also give a set of fast, separable filters that generalizes Gaussian filtering to vectors. Similarly, we study associated splines in Section 4, and show how the coupling can improve motion interpolation of points. These splines generalize Laplacian splines, such as the thin plate splines, to vectors. In Section 5, we finally propose to merge filters and splines for intensity- and geometric-feature based registration.

2 Isotropic Energies

Throughout this article, we note $\mathbb{R}$ the set of real numbers. A d-dimensional vector field $\mathbf{f}$ is an application from $\mathbb{R}^d$ to $\mathbb{R}^d$. If $\mathbf{f}$ is a d-dimensional vector field, we note f_i its i-th coordinate, so that $\mathbf{f} = (f_1, ..., f_d)^T$.

If f is a function from $\mathbb{R}^d$ to $\mathbb{R}$, we note $\partial_i f$ its partial derivative with respect to its i-th parameter. By $\partial_{ijk\cdots}$ we note the composition of the partial derivatives operators, i.e. $\partial_{ijk\cdots} = \partial_i \partial_j \partial_k \cdots$.

In the following, we will often use a simplified Einstein summation convention, where every index repeated twice in a product is implicitly summed all over its range (e.g. if $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$, then $f_i f_i$ means $f_1^2 + f_2^2$).

This article relies on the use of the following family of single-order quadratic energies:

$$E_n(\mathbf{f}) = \partial_{i_1 \ldots i_n} f_{i_{n+1}} \partial_{i_1 \ldots i_n} f_{i_{n+1}} + \kappa . \partial_{i_1 \ldots i_n} f_{i_{n+1}} \partial_{i_{n+1} i_2 \ldots i_n} f_{i_1} \tag{1}$$

where $\mathbf{f}$ is the vector field we want to regularize. This energy E_n depends on the n-th order partial derivatives of $\mathbf{f}$.

The isotropy of E_n, Eq. (1), is a consequence of the fact that products of Kronecker tensors are isotropic. It can be easily verified: according to the chain rule, if $R = (R_{ij})$ is an orthogonal transformation, $\partial_{k_1 \dots k_n} f \circ R = R_{i_1 k_1} \dots R_{i_n k_n} (\partial_{i_1 \dots i_n} f) \circ R$. Then, if σ is one-to-one mapping of the numbers 1 to n onto themselves,

$$\begin{aligned} \partial_{k_1 \dots k_n} f \circ R . \partial_{k_{\sigma(1)} \dots k_{\sigma(n)}} f \circ R &= R_{i_1 k_1} \dots R_{i_n k_n} (\partial_{i_1 \dots i_n} f) \\ &\quad \circ R . R_{j_1 k_{\sigma(1)}} \dots R_{j_n k_{\sigma(n)}} (\partial_{j_1 \dots j_n} f) \circ R \\ &= (\partial_{k_1 \dots k_n} f) \circ R . (\partial_{k_{\sigma(1)} \dots k_{\sigma(n)}} f) \circ R \end{aligned}$$

because $R_{i_1 k_1} \dots R_{i_n k_n} R_{j_1 k_\sigma(1)} \dots R_{j_n k_\sigma(n)}$ is one iff $j_m = i_{\sigma(m)}$ $\forall m = 1 \dots d$ (otherwise, it is zero).

The quadratic energy E_n, Eq. (1), is designed for regularization. Therefore, it should have a lower bound, reached for deformations considered to be perfectly smooth, e.g. the identity. For quadratic forms, this condition is equivalent to positiveness. We should therefore calculate the eigenvalues of E_n and deduce positiveness conditions.

Its eigentensors and associated eigenvalues are:

Eigentensor	Eigenvalue
$E^{(i_1, \dots i_{n+1})} + E^{(i_{n+1}, i_2, \dots i_n, i_1)}$	$1 + \kappa$
$E^{(i_1, \dots i_{n+1})} - E^{(i_{n+1}, i_2, \dots i_n, i_1)}$	$1 - \kappa$
$E^{(k, i_2, \dots i_n, k)}$	$1 - \kappa$

where $E^{(i_1, \dots i_{n+1})}$ is the tensor whose only non-zero element is $E^{(i_1, \dots i_{n+1})}_{i_1, \dots i_{n+1}} = 1$. One can verify that all these d^n tensors are orthogonal. Therefore, there exists no additional eigenvalue. The positiveness conditions is $\kappa \geq -1$.

Single-order energies can finally be summed to create multi-order energies, say $R = \sum_{n=1}^{\infty} \lambda_n E_n$.

3 Isotropic Convolution Filters for Vector Fields

Smoothing filters can be deduced from quadratic regularization energies R as the closed-form solution of the homogeneous dense approximation problem. This problem consists in approximating a vector field $\mathbf{g}$, supposed noisy and irregular, by a smooth vector field $\mathbf{f}$ by minimizing

$$E(\mathbf{f}) = \int ||\mathbf{f} - \mathbf{g}||^2 + R(\mathbf{f}) \tag{2}$$

A closed-form solution can be obtained in the Fourier domain by using functional derivatives of the energy. We thus have to calculate the derivative of R.

3.1 Functional Derivatives

The functional derivative of E_n w.r.t $\mathbf{f}$ is $2(-1)^n \Delta^n \mathbf{f} + 2\kappa \Delta^{n-1} \nabla \nabla^T \mathbf{f}$ where $\nabla = (\partial_1, \dots, \partial_d)^T$ is the usual partial derivative operator, and $\Delta = \nabla^T \nabla$.Id is the

Laplacian operator. Now, the functional derivative of R is $2\sum_{n=1}^{\infty}(-1)^n\lambda_n\Delta^{n-1} \times (\Delta + \kappa\nabla\nabla^T)\mathbf{f}$ and the optima of Eq. (2) now yields

$$\sum_{n=0}^{\infty}(-1)^n\lambda_n\Delta^{n-1}(\Delta + \kappa\nabla\nabla^T)\mathbf{f} = \mathbf{g} - \mathbf{f} \tag{3}$$

3.2 In the Fourier Domain

We rewrite Eq. (3) in the Fourier domain, in order to transform differential operators into multiplicative coefficients.

Noting $\mathbf{x} = (x_1, ..., x_d)$ and $\mathbf{w} = (w_1, ..., w_d)$ the canonical coordinates of resp. the spatial and the Fourier domain, and $\hat{\cdot}$ the Fourier operator, Eq. (3) transforms in the Fourier domain into

$$\underbrace{\sum_{n=1}^{\infty}\left[\left(1 + \lambda_n(\mathbf{w}^T\mathbf{w})^n\right)\mathrm{Id} + \lambda_n\kappa_n(\mathbf{w}^T\mathbf{w})^{n-1}(\mathbf{w}\mathbf{w}^T)\right]}_{M}\hat{\mathbf{f}} = \hat{\mathbf{g}}$$

We then need to invert matrix M to solve our problem in three steps, in the Fourier domain: 1), Compute the Fourier transform of $\mathbf{g}$, the vector field to be smoothed; 2), Multiply the result by M^{-1}; 3), Compute the inverse Fourier transform of the result, which gives the solution. We will now review particular cases of such filters for single-order regularization energies.

3.3 First-Order Isotropic Filters

In the first-order case, matrix M becomes $M_1 = \left[\left(1 + \lambda\mathbf{w}^T\mathbf{w}\right)\mathrm{Id} + \lambda\kappa\mathbf{w}\mathbf{w}^T\right]$. The closed-form of its inverse is

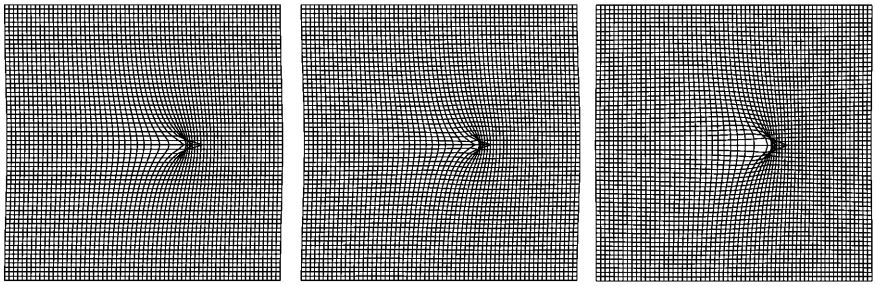
$$M_1^{-1} = \frac{1}{1 + \lambda\mathbf{w}^T\mathbf{w}}\left[\mathrm{Id} - \frac{\lambda\kappa}{1 + \lambda(1+\kappa)\mathbf{w}^T\mathbf{w}}\mathbf{w}\mathbf{w}^T\right]$$

We show three examples of first order isotropic filters, depending on the sign of the cross-effect parameter κ, which act similarly as the Poisson ratio ν of the theory of elasticity. Here, the vector field $\mathbf{g}$ to be smoothed is a simple impulse in the horizontal direction. The results shown in Fig. 1 can be considered as the impulse response of their respective filter.

The first filter, Fig. 1a, does not present any cross-effect ($\kappa = 0$) and corresponds to a membrane model. Without cross effects, horizontal lines stay straight: there is no motion along the vertical axis because the input impulse itself has no vertical component.

The second filter, Fig. 1b, presents cross-effects ($\kappa > 0$); the tissue gets closer to an incompressible model and the material is less deformed: this can be interesting for registration purposes, for example if the organs to be registered are nearly incompressible, such as the brain.

The last filter, Fig. 1c, also has cross-effects ($\kappa < 0$), but its behavior is somewhat counter-intuitive as the impulse tends to inflate the material behind it.



a. $\lambda = 5$, $\kappa = 0$ b. $\lambda = 5$, $\kappa = 0.5$ c. $\lambda = 5$, $\kappa = -0.5$

Fig. 1. Impulse response of three different first order isotropic filters.

Although some rare materials do have this kind of behavior, e.g. some particular foams [6], the choice of a negative κ for registration has more to do with a prior knowledge on the displacement (e.g. inflation or shrinkage).

For any value of λ or κ, there exists a discontinuity in the derivative of the impulse response, at the tip of the sharp peak. In non-rigid registration, external forces are dense, and applied at every pixel of the image, therefore such peaks may appear frequently in the image. This is particularly annoying if a further processing uses a differential analysis of the computed displacement field, e.g. a Jacobian based segmentation as in [7]. For this purpose, we need filters of order higher than these linear elastic filters.

3.4 Second-Order Isotropic Filters

In the second-order case, matrix M becomes

$$M_2 = \left[\left(1 + \lambda(\mathbf{w}^T\mathbf{w})^2\right) \mathrm{Id} + \lambda\kappa(\mathbf{w}^T\mathbf{w})(\mathbf{w}\mathbf{w}^T)\right]$$

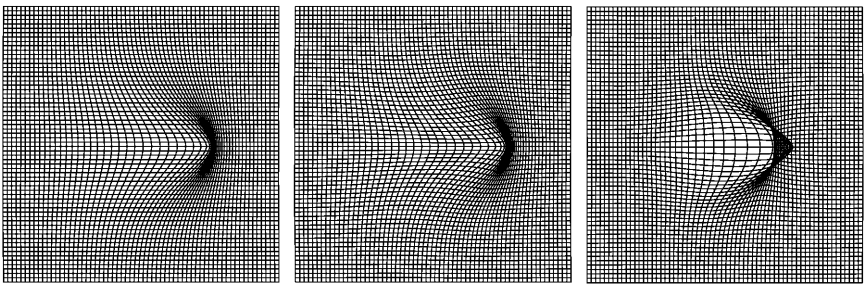
The closed-form of its inverse is

$$M_2^{-1} = \frac{1}{1 + \lambda(\mathbf{w}^T\mathbf{w})^2}\left[\mathrm{Id} - \frac{\lambda\kappa}{1 + \lambda(1+\kappa)(\mathbf{w}^T\mathbf{w})^2}\mathbf{w}\mathbf{w}^T\right]$$

In Fig. 2, we present the impulse response of the second order filter for three different values of κ, one without any cross-effect ($\kappa = 0$) corresponding to a thin-plate model, and two others with positive and negative cross-effects. The impact of κ on the filter is similar as in the first order case, i.e. that of a Poisson ratio. However, impulse responses are much smoother, which makes differential analysis and one-to-one mapping easier.

3.5 Separable Isotropic Filters

Previous vectorial convolution filters are applied in the Fourier domain, which is generally the faster way to proceed, especially in 3D and for large kernels. However, the computation can be even faster in the real domain, if the kernel is separable.



a. $\lambda = 2$, $\kappa = 0$ b. $\lambda = 2$, $\kappa = 0.6$ c.$\lambda = 2$, $\kappa = -0.6$

Fig. 2. Impulse response of three different second order isotropic filters.

Definition 1 (Isotropic Filters) *A scalar convolution kernel* $f(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}$ *is isotropic if* $f(R^T\mathbf{x}) = f(\mathbf{x})$, $\forall R \in O(d)$. *A vector convolution kernel* $F(\mathbf{x}) = F(x_1, ..., x_d) : \mathbb{R}^d \to \mathcal{M}_d$ *is isotropic if* $RF(R^T\mathbf{x})R^T = F(\mathbf{x})$, $\forall R \in O(d)$.

Definition 2 (Separable Filters) *A scalar convolution kernel* $f(\mathbf{x}) = f(x_1, ..., x_d) : \mathbb{R}^d \to \mathbb{R}$ *is separable if there exists d functions* f_k *such that* $f(\mathbf{x}) = f_1(x_1)f_2(x_2)...f_d(x_d) = \prod_{k=1}^{d} f_k(x_k)$. *A vector convolution kernel* $F(\mathbf{x}) = F(x_1, ..., x_d) : \mathbb{R}^d \to \mathcal{M}_d$ *is separable if each of its element* $F_{i,j}$ *is separable.*

Note that there are other possible definitions of separability for vectorial filters. The property of separability is very interesting from a numerical point of view, because an n-dimensional convolutions then boils down to a sequence of 1-D convolutions, which can be implemented efficiently for instance using recursive filtering.

The choice of separable filters is drastically reduced if we also impose the isotropy property. For scalars, it is known that the only isotropic separable kernels are the family of Gaussians [5]. However, to the best of our knowledge, there is no similar theorem for vector filters. We propose the following result, easy to verify:

Proposition 1 *The vector filter* $G_{\sigma,\kappa}$ *defined by*

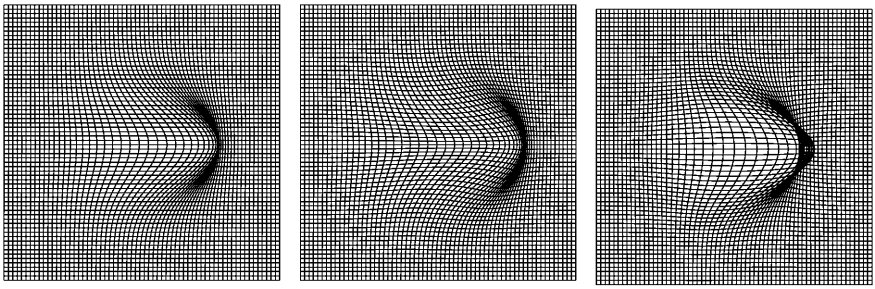
$$G_{\sigma,\kappa}(\mathbf{x}) = \frac{1}{(\sigma\sqrt{2\pi})^d(1+\kappa)}(Id_d + \frac{\kappa}{\sigma^2}\mathbf{x}\mathbf{x}^T)e^{-\frac{\mathbf{x}^T\mathbf{x}}{2\sigma^2}} \tag{4}$$

is separable and isotropic.

When κ is set to zero, the matrix $G_{\sigma,0}$ is diagonal and we obtain a classical Gaussian filtering, independently on each component of the vector field.

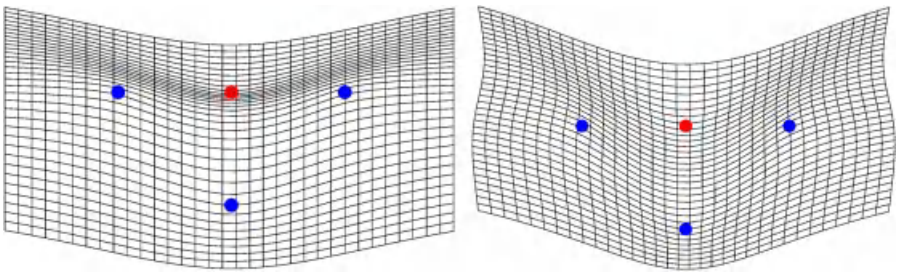
There exist separable isotropic filters that are not part of this previous family, as this filter in 2D:

$$\frac{1}{(\sigma\sqrt{2\pi})^2(1-\kappa)}\begin{pmatrix} 1-\kappa y^2/\sigma^2 & \kappa xy/\sigma^2 \\ \kappa xy/\sigma^2 & 1-\kappa x^2/\sigma^2 \end{pmatrix} e^{-\frac{x^2+y^2}{2\sigma^2}}$$



a. $\sigma = 8$, $\kappa = 0$ b. $\sigma = 8$, $\kappa = 0.12$ c. $\sigma = 8$, $\kappa = -0.12$

Fig. 3. Impulse response of three different separable isotropic vector filters $G_{\sigma,\kappa}$, Eq. (4) generalizing the classical scalar Gaussian filtering to vectors.



a. Using TPS b. Using vectorial TPS

Fig. 4. Interpolation of displacements using, left, standard scalar thin-plate splines, and right, vectorial thin-plate splines.

However we could not find counter examples in higher dimension. The filters $G_{\sigma,\kappa}$ given by (4) might be the only separable isotropic filters in dimension greater than or equal to 3, but this has yet to be proved.

Figure 4 shows the impulse responses of the separable isotropic filter (4) for three different values of parameter κ. Their behavior is somewhat similar to second order filters, Fig. 2, although here the impulse response is even smoother, as the underlying energy is of infinite order.

4 Isotropic Vectorial Splines

In the same way filters have been deduced from regularization energies in the previous section, splines can be designed *via* the sparse vector interpolation and approximation problems, which are a straightforward extension of the scalar ones [4]. Given two sets of points $\mathbf{x}_i$ and $\mathbf{g}_i$ of $\mathbb{R}^d$, the vector interpolation problem consists in finding a vector field $\mathbf{f} : \mathbb{R}^d \to \mathbb{R}^d$ which minimizes a regularization energy $R(\mathbf{f})$ under the constraint that $\mathbf{f}(\mathbf{x}_i) = \mathbf{g}_i$. The vector approximation problem consists in seeking

$$\mathbf{f}^* = \arg\min_{\mathbf{f}} \sum_i ||\mathbf{g}_i - \mathbf{f}(\mathbf{x}_i)||^2 + \lambda R(\mathbf{f})$$

When $\lambda \to 0$, the approximation solution $\mathbf{f}^*$ tends toward the interpolation solution. More elaborate approximation problems are sometimes useful, where for example the isotropic distance $||\mathbf{g}_i - \mathbf{f}(\mathbf{x}_i)||^2$ is replaced by an anisotropic distance depending on the index i [9].

4.1 Vectorial Laplacian Splines

For vector interpolation and approximation, the most common solution in the field of non-rigid registration consists in interpolating or approximating every component independently with isotropic scalar kernel, such as Laplacian splines or radial basis functions. A notable exception is the elastic body spline of [3], which is based on linear elasticity. Unfortunately, the exact interpolation kernel of linear elasticity is ill-defined as it tends toward infinity at its center. Therefore, there is a need of higher-order splines having tunable cross-effects.

For vector interpolation or approximation problems, we propose to generalize Laplacian splines using the energy given by Eq. (1). As for the scalar case [4], if $n > d/2$, we note S_n the solution of

$$(-1)^n \left[\Delta^n S_n + \kappa \Delta^{n-1} \nabla \nabla^T S_n\right] = \delta \mathrm{Id} \tag{5}$$

then the solution $\mathbf{f}$ of the vector interpolation or approximation problem has the following form $\mathbf{f}(\mathbf{x}) = \mathbf{p}(\mathbf{x}) + \sum_i S_n(\mathbf{x} - \mathbf{x}_i)\boldsymbol{\alpha}_i$, where $\mathbf{p}$ is a polynomial such that $E_n(\mathbf{p}) = 0$, and $\boldsymbol{\alpha}_i \in \mathbb{R}^d$ are coefficients that can be found by solving a set of linear equations.

It is possible to get closed-form formulas for S_n by getting back to the scalar case. In the Fourier domain, the PDE (5) becomes $\left[(\mathbf{w}^T\mathbf{w})^n \mathrm{Id} + \kappa(\mathbf{w}^T\mathbf{w})^{n-1}\mathbf{w}\mathbf{w}^T\right]\hat{S}_n = \mathrm{Id}$ yielding

$$\hat{S}_n = \frac{1}{(\mathbf{w}^T\mathbf{w})^n}\mathrm{Id} - \frac{\kappa}{(1+\kappa)(\mathbf{w}^T\mathbf{w})^{n+1}}\mathbf{w}\mathbf{w}^T$$

The Fourier transform of the scalar Laplacian spline s_n is precisely $\frac{1}{(\mathbf{w}^T\mathbf{w})^n}$. Thus, we can express S_n as a function of s_n and s_{n+1}:

$$S_n = s_n \mathrm{Id} + \frac{\kappa}{1+\kappa}\mathcal{H}s_{n+1}$$

where $\mathcal{H}s_{n+1}$ is the Hessian matrix of s_{n+1}. The closed-forms s_n are provided by [4]: $|x|^{2d-n}\log|x|$ if d is even, $|x|^{2d-n}$ if d is odd.

Example: 2D Vectorial Thin Plate Spline Interpolation. Up to a multiplicative coefficient, the 2-D second order Laplacian spline (or thin plate spline) is $s_2(x,y) = r^2 \ln r^2$, and the 2-D third-order Laplacian spline is $s_3(x,y) = r^4 \ln r^2$, where $r^2 = x^2 + y^2$. One can thus calculate $\mathcal{H}s_3$, and find the close form formula of S_2 in 2D: $(S_2)_{1,1} = r^2 \ln r^2 + \frac{\kappa}{1+\kappa}\left[2(3x^2+y^2)\ln r^2 + 7x^2 + y^2\right]$, $(S_2)_{1,2} = 2\frac{\kappa}{1+\kappa}xy(2\ln r^2 + 3)$.

In Fig. 7, we compare the results of the interpolation of displacements using scalar and vectorial thin plate splines. In the original position, the four points were placed at each corner of a square. The upper point has then been forced to move to the center of the square. The scalar thin-plate spline interpolation applied on each component independently, Fig. 7a, do not present any horizontal displacement, and thus vertical lines remains straight. It possesses a strong accumulation of matter just under the point that has been moved. The vectorial Laplacian spline interpolation, Fig. 7b, is more realistic thanks to a better distribution of the resulting displacements into both vertical and horizontal components; the previous accumulation of matter has disappear.

5 Merging Filters and Splines

In this section, we merge the two previous problems of dense (Section 3) and sparse (Section 4) vector approximation: we are searching for a vector field $\mathbf{f} : \mathbb{R}^d \to \mathbb{R}^d$ that approximates both a dense vector field $\mathbf{g}_1$, and a set of discrete pairings $\mathbf{g}_2$. The energy to minimize is

$$E(\mathbf{f}) = \int ||\mathbf{f} - \mathbf{g}_1||^2 + \gamma \sum_{i=1\ldots p} ||\mathbf{f}(\mathbf{x}_i) - \mathbf{g}_2(\mathbf{x}_i)||^2 + R(\mathbf{f})$$

It can be shown [1] that the optimal solution $\mathbf{f}$ is a linear combination of a smoothed vector field and a sum of splines $\mathbf{f}(\mathbf{x}) = K * \mathbf{g}_1(\mathbf{x}) + \sum_i K(\mathbf{x} - \mathbf{x}_i)\boldsymbol{\alpha}_i$, where K is the smoothing kernel associated to R, as those found in Section 3, and $\boldsymbol{\alpha}_i \in \mathbb{R}^d$ are coefficients found by solving a set of linear equations. Note that the same kernel is used both as a smoothing kernel to smooth C_1, and as a spline to approximate the sparse correspondences C_2. Contrary to the vectorial Laplacian splines of Section 4.1, these splines are bounded, decrease toward zero at infinity, and exist in any space dimension.

This formulation turns out to be very useful in the context of non-rigid registration, when one wants to add sparse geometric constraints to an intensity based algorithm, because the intensity similarity measure gives dense pairings $\mathbf{g}_1$, while geometric constraint gives sparse pairings $\mathbf{g}_2$.

We successfully applied this technique to the difficult problem of intersubject brain registration [2]. Because of the high variability of the cortex topology between different subjects, it is difficult to obtain valid registration results using the intensity alone. On the other hand, geometric features provided for instance by a set of sulcal lines extracted on the cortex [8], are too sparse to provide an accurate displacement field everywhere. The idea developed in [2] is to combine intensity-based and feature-based registration with the above presented technique.

6 Conclusion

In this paper we introduced some new techniques for regularizing vector fields. We introduced isotropic quadratic energies and deduced vector filters and splines

to approximate respectively dense and sparse vector fields. We also introduced a separable vector filter that generalizes Gaussian filtering to vectors and enables a particularly efficient smoothing, using recursive filtering. Finally, we combined both sparse and dense approximation problems, whose resulting closed-form solution can be applied successfully to real problems.

The original feature of vector regularization is the possibility to have cross-effects between coordinates, which is not possible using standard scalar regularization on each component separately. This new parameter makes it possible to tune more finely the solution to our problem, in the context of non-rigid registration for example, depending on our prior knowledge. Future quantitative analysis should demonstrate improvement of the motion recovery using the models proposed in this paper depending on the nature of the deformed material.

References

1. P. Cachier, E. Bardinet, D. Dormont, X. Pennec, and N. Ayache. Iconic Feature Based Nonrigid Registration: The PASHA Algorithm. *CVIU — Special Issue on Nonrigid Registration*, 2003. In Press.
2. P. Cachier, J.-F. Mangin, X. Pennec, D. Rivière, D. Papadopoulos-Orfanos, J. Régis, and N. Ayache. Multisubject Non-Rigid Registration of Brain MRI using Intensity and Geometric Features. In *Proc. of MICCAI'01*, volume 2208 of *LNCS*, pages 734 – 742, Utrecht, the Netherlands, October 2001.
3. M. H. Davis, A. Khotanzad, D. P. Flamig, and S. E. Harms. A Physics-Based Coordinate Transformation for 3D Image Matching. *IEEE Trans. on Medical Imaging*, 16(3):317 – 328, June 1997.
4. J. Duchon. Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces. *RAIRO Analyse Numérique*, 10(12):5 – 12, December 1976.
5. P. Kannappan and P. K. Sahoo. Rotation Invariant Separable Functions are Gaussian. *SIAM J. on Math. Analysis*, 23(5):1342 – 1351, September 1992.
6. R. Lakes. Foam Structures with a negative Poisson's ratio. *Science*, 235(4792):1038 – 1040, February 1987.
7. D. Rey, G. Subsol, H. Delingette, and N. Ayache. Automatic Detection and Segmentation of Evolving Processes in 3D Medical Images: Application to Multiple Sclerosis. In *Proc. of IPMI'99*, volume 1613 of *LNCS*, pages 154–167, Visegrád, Hungary, June 1999.
8. D. Rivière, J.-F. Mangin, D. Papadopoulos, J.-M. Martinez, V. Frouin, and J. Régis. Automatic Recognition of Cortical Sulci using a Congregation of Neural Networks. In *Proc. of MICCAI'00*, volume 1935 of *LNCS*, pages 40 – 49, Pittsburgh, USA, October 2000. Springer.
9. K. Rohr, M. Fornefett, and H. S. Stiehl. Approximating Thin-Plate Splines for Elastic Registration: Integration of Landmark Errors and Orientation Attributes. In *Proc. of IPMI'99*, volume 1613 of *LNCS*, pages 252 – 265, Visegrád, Hungary, June/July 1999. Springer.

Free-Form Registration Using Mutual Information and Curvature Regularization

E. D'Agostino[1], J. Modersitzki[2], F. Maes[1,⋆], D. Vandermeulen[1], B. Fischer[2], and P. Suetens[1]

[1] K.U.Leuven, Faculties of Medicine and Engineering
Medical Image Computing (Radiology - ESAT/PSI)
UZ Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium
{Emiliano.DAgostino,Frederik.Maes}@uz.kuleuven.ac.be
[2] Universität Lübeck, Institut für Mathematik
Wallstraße 40, 23560 Lübeck, Germany
{modersitzki,fischer}@math.uni-luebeck.de

Abstract. In this paper we present a novel 3-D free-form non-rigid registration algorithm which combines the mutual information similarity measure with a particular curvature based regularizer, which has been demonstrated to produce very satisfactory results in conjunction with the sum of squared differences distance measure. The method is evaluated for inter-subject MR brain image registration using simulated deformations and compared with a scheme that applies the same similarity measure but with a viscous fluid regularizer.

1 Introduction

Non-rigid image registration of two three-dimensional (3-D) medical image volumes involves finding the 3-D vector field of 3-D displacements that maps each point in the reference image onto its anatomically corresponding point in the template image, such that the template image can be geometrically deformed or warped to exactly match the reference image. Typical applications include atlas construction, atlas-based segmentation, shape analysis or motion estimation. Recovering the deformation field from the image data itself requires specification of a proper similarity measure for assessing the quality of the match and of a suitable regularization scheme to exclude non-realistic deformation and to assure that the problem is mathematically well-posed.

Maximization of mutual information (MMI) of corresponding voxel intensities has been demonstrated to be highly succesful for affine image registration [7] and several approaches have been proposed to extend the MMI criterion to non-rigid matching. These differ in the way mutual information (MI) is computed when varying the registration parameters and in the regularization constraints that are imposed on the deformation field. While ideally regularization should

⋆ F. Maes is Postdoctoral Fellow of the Fund for Scientific Research - Flanders (FWO-Vlaanderen, Belgium).

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 11–20, 2003.

incorporate biomechanical or statistical models of tissue deformation, in practice most regularization schemes for non-rigid registration are purely mathematical and can be categorized as either implicit or explicit. Implicit schemes represent the deformation field using basis functions with built-in smoothness, e.g. B-splines [10], such that the registration solution is constrained to be within a particular class of transformations. Explicit schemes on the other hand allow for free-form deformations of individual voxels, but penalize displacements that violate local smoothness. The main advantage of explicit regularization is the possibility of incorporating physical properties of the objects to be registered, like for example elasticity. Moreover, this approach is more flexible, in the sense that the set of transformations that can be recovered, is in general much larger. In addition, the numerical treatment of the explicitly regularized registration problems leads to systems of partial differential equations for which fast, stable, and efficient solvers are available.

Recently, Hermosillo *et al.* [6] and D'Agostino *et al.* [1] constructed a voxel-wise force field that allows to drive free-form registration such as to maximize MI, by deriving the gradient of MI with respect to single voxel displacements using a continuous and differentiable representation of the joint intensity histogram based on Parzen estimation. While [6] and [1] combined the MMI force-field with an elastic and a viscous fluid regularizer respectively, in this paper we apply the MMI criterion in conjunction with the curvature based regularizer presented by Fischer and Modersitzki [4] . In contrast to other regularizers, affine linear transformations are not penalized by the curvature regularizer, such that non-rigid registration is less sensitive to non-optimal affine pre-registration or may even be applied without prior affine registration. The purpose of this work is two-fold: firstly, we demonstrate that the force field derived in [6,1] can be applied succesfully to drive MMI using different regularization kernels, such that various regularizers can be investigated for a particular application in conjunction with the same MI similarity measure; secondly, we present some initial results comparing the performance of the viscous fluid and curvature based regularizers for inter-subject MR brain image registration, both allowing large deformations. Our results indicate that acceptable and comparable registration results can be obtained using different regularization schemes and different numerical solvers, such that selecting an optimal regularizer is non-trivial.

The paper is organized as follows. In Section 2 we describe the curvature based regularization scheme and the numerical solver proposed in [4] and discuss some implementation issues. In Section 3 we investigate the performance of the curvature regularizer in function of its parameters and present a quantitative comparison with the viscous fluid regularizer of [1] for recovering simulated deformations of MR brain images.

2 Method

2.1 Similarity Measure

Given two image volumes R and T defined on a domain $\Omega \subset \mathbb{R}^3$ with $T(x)$ denoting the intensity of T at $x \in \Omega$, the purpose of non-rigid image registration

is to find the deformation field $u : \mathbb{R}^3 \to \mathbb{R}^3$ that maps points in R onto their corresponding points in T, such that the deformed template $T_u = T(x - u(x))$, becomes similar to R. Mutual information measures the similarity between R and T_u by the Kullback-Leibler distance

$$\mathrm{MI}[R, T; u] = \int_{\mathbb{R}^2} p^{R,T_u} \log \frac{p^{R,T_u}(g_1, g_2)}{p^R(g_1)p^{T_u}(g_2)} \; d(g_1, g_2), \tag{1}$$

with $p^{R,T_u}(g_1, g_2)$ the joint intensity density of intensities g_1 in R and g_2 in T_u with marginal densities $p^R(g_1)$ and $p^{T_u}(g_2)$ respectively. Following the approach of [1], we estimate the density p^{R,T_u} by the Parzen-window density estimator $\hat{p}^{R,T}$ based on a sample Ω_d of Ω using a Gaussian Parzen-window function $\Psi_\sigma : \mathbb{R}^2 \to \mathbb{R}$ with width σ whose choice is discussed below:

$$\Psi_\sigma(g_1, g_2) = \frac{1}{2\pi\sigma} \exp\big(-\frac{g_1^2 + g_2^2}{2\sigma}\big),$$

$$p^{R,T}(g_1, g_2) \approx \hat{p}^{R,T}(g_1, g_2) := \frac{1}{\#\Omega_d} \sum_{x \in \Omega_d} \Psi_\sigma(g_1 - R(x), g_2 - T(x)), \tag{2}$$

$$p^R(g_1) \approx \hat{p}^R(g_1) := \int_{\mathbb{R}} \hat{p}^{R,T}(g_1, g_2) \; dg_2,$$

$$p^T(g_2) \approx \hat{p}^T(g_2) := \int_{\mathbb{R}} \hat{p}^{R,T}(g_1, g_2) \; dg_1,$$

such that

$$\mathrm{MI}[R, T; u] \approx \hat{\mathrm{MI}}[R, T; u] := \int_{\mathbb{R}^2} \hat{p}^{R,T_u} \log \frac{\hat{p}^{R,T_u}}{\hat{p}^R \hat{p}^{T_u}} \; d(g_1, g_2) \tag{3}$$

2.2 The Regularizer

Different cost functionals $\mathcal{S}$ for regularization of free-form deformations have been proposed in the literature and have been evaluated by Modersitzki [8] for unimodal image registration in conjunction with the sum of squared differences similarity measure. In this paper, we focus on the so-called *curvature* regularizer $\mathcal{S}^{\mathrm{curv}}$ introduced in [4] :

$$\mathcal{S}^{\mathrm{curv}}[u] = \sum_{\ell=1}^{3} \int_\Omega (\Delta u_\ell)^2 \; dx. \tag{4}$$

Apart from its smoothness, the main advantage of this regularizer is its ability to automatically correct for affine linear deformations. This is due to the fact that the smoother is based purely on second order derivatives which do not penalize affine linear transformations, i.e.,

$$\mathcal{S}^{\mathrm{curv}}[Bx + c] = 0 \quad \text{for all} \quad B \in \mathbb{R}^{3\times 3}, \; c \in \mathbb{R}^3.$$

Thus, in view of the fact that an affine linear pre-registration may be suboptimal, the curvature regularizer allows for an automatic correction of the affine linear parts. Other regularizers, such as the elastic, fluid, or diffusion regularizers, do not have this feature. It has been validated that the curvature regularizer is much less sensitive to the initial position of the images to be registered [4] .

Combining the MI similarity measure and the regularization cost in a single functional, the registration problem consists of finding the deformation u which minimizes the joint criterion

$$\mathcal{J}[u] := -\hat{\mathrm{MI}}[R, T; u] + \alpha \mathcal{S}^{\mathrm{curv}}[u], \tag{5}$$

where the parameter $\alpha > 0$ controls the strength of the regularization versus the similarity of the images. The role of α is investigated in Section 3 below.

2.3 Optimization

To compute a solution for this minimization problem, we use a steepest descent method considering that the first variation of the combined functional $\mathcal{J}$ vanishes at the optimum. To this end, one has to compute the Gâteaux-derivatives of the participating functionals.

For $u \in C^0(\Omega)^3$ and for each perturbation $v \in C^0(\Omega)^3$, the Gâteaux-derivative of $\hat{\mathrm{MI}}$ as derived by Hermosillo *et al.* [6] is given by

$$d\hat{\mathrm{MI}}[R, T; u; v] = \frac{1}{\#\Omega_d} \sum_{x \in \Omega_d} \left\langle f^{R,T}(x, u(x)), v(x) \right\rangle_{\mathbb{R}^3} \tag{6}$$

where

$$f^{R,T}(x, u(x)) = [\Psi_\sigma * \partial_{g_2} L^{R,T_u}](R(x), T_u(x)) \cdot \nabla T(x - u(x)), \tag{7}$$

$$L^{R,T_u}(g_1, g_2) := 1 + \hat{p}^{R,T_u} \log \frac{\hat{p}^{R,T_u}}{\hat{p}^R \hat{p}^{T_u}},$$

and with $*$ being the convolution operator:

$$[p * q](z_1, z_2) := \int_{\mathbb{R}^2} p(z_1 - g_1, z_2 - g_2) q(g_1, g_2)\ d(g_1, g_2).$$

For $u \in C^4(\mathbb{R}^3)^3$ and for each perturbation $v \in C^4(\mathbb{R}^3)^3$, the Gâteaux-derivative of $\mathcal{S}^{\mathrm{curv}}$ as derived by Fischer and Modersitzki [4] is given by

$$d\mathcal{S}^{\mathrm{curv}}[u; v] = \int_\Omega \left\langle \Delta^2 u, v \right\rangle_{\mathbb{R}^3}\ dx + \sum_{\ell=1}^{3} \int_{\partial\Omega} \Delta u_\ell \left\langle \nabla v_\ell, \boldsymbol{n} \right\rangle_{\mathbb{R}^3} - v_\ell \left\langle \nabla \Delta u_\ell, \boldsymbol{n} \right\rangle_{\mathbb{R}^3}\ dx, \tag{8}$$

where $\boldsymbol{n}$ denotes the outer normal unit vector on the boundary $\partial\Omega$. Imposing the explicit boundary conditions on u and v,

$$\nabla u_\ell = \nabla \Delta u_\ell = 0 \quad \text{on} \quad \partial\Omega, \quad \ell = 1, 2, 3, \tag{9}$$

the boundary integrals in the expression for $d\mathcal{S}^{\mathrm{curv}}[u; v]$ vanish and we obtain

$$d\mathcal{S}^{\text{curv}}[u;v] = \int_{\Omega} \left\langle \Delta^2 u, v \right\rangle_{\mathbb{R}^d} \, dx. \tag{10}$$

In accordance with the calculus of variations, a function $u \in C^4(\mathbb{R}^d)^d$ which minimizes the joint functional (5) has to satisfy the Euler-Lagrange equation

$$\alpha \Delta^2 u + f^{R,T}(x, u(x)) = 0 \quad \text{for all} \quad x \in \Omega \tag{11}$$

subject to the boundary condition (9), where $f^{R,T}$ is given by (7).

The Euler-Lagrange equation (11) is a fourth-order non-linear partial differential equation (PDE). It is known as the bipotential or biharmonic equation and is well understood; see, e.g. [5] . For applications in mechanics, this equation describes the displacement of a thin plate subject to the load $f^{R,T}$. In the context of image registration, the quantity $f^{R,T}$ may be seen as a force field which drives the template towards the reference subject to the physical constraints imposed by the biharmonic operator and the boundary conditions. Also, it should be noted that there is a close connection between the curvature regularizer $\mathcal{S}^{\text{curv}}$ (4) and the functional which is minimized by the so-called *thin-plate-splines*, cf., e.g., [9]. The GÂTEAUX-derivatives of both functionals share the same main part but differ slightly in their boundary integrals. However, this close connection offers an alternative interpretation of the minimizer of (5). The final displacement is as smooth as thin-plate-splines but its shape is not determined by any user supplied landmarks, merely it is automatically designed by the forces imposed by the MI measure.

2.4 Implementation

The MI force field (7) depends on the width σ of the Gaussian Parzen-window kernel. A proper value for σ is determined as $\sigma = \max\{\sigma^R, \sigma^T\}$, with σ^R and σ^T estimated for R and T separately as the kernel widths that maximize the log-likelihood of the marginal densities $p_\sigma^R(g)$ and $p_\sigma^T(g)$ respectively, constructed by the Parzen-window estimator using a leave-one-out approach:

$$\begin{aligned}
p^R(g) &= c_h \cdot \#\{x : \ R(x) = g\}, \\
p_\sigma^R(g) &= \frac{c_\sigma}{\#\Omega(g)} \sum_{x \in \Omega(g)} \psi_\sigma(g - R(x)) = \frac{c_\sigma}{c_h \cdot \#\Omega(g)} \sum_{q \neq g} h^R(q) \cdot \psi_\sigma(g - q), \\
\sigma^R &= \arg\max_{\sigma} \sum_{g \in G} p^R(g) \log p_\sigma^R(g).
\end{aligned}$$

where $\Omega(g) := \{x \in \Omega_d : \ R(x) \neq 0 \wedge R(x) \neq g\}$, ψ_σ is a one-dimensional Gaussian, and the constants c_h and c_σ are chosen such that $\sum_{g \in G} h^B(g) = 1$ and $\sum_{g \in G} h_\sigma^B(g) = 1$, respectively, and likewise for σ^T. This approach is identical to the one described in [1].

We solve the Euler-Lagrange equation (11) by applying a semi-implicit discrete time-marching algorithm

$$\frac{u^{(k+1)}(x) - u^{(k)}(x)}{\tau} + \alpha \Delta^2 u^{(k+1)}(x) = -f^{R,T}(x, u^{(k)}(x)) \quad \text{for all} \quad x \in \Omega,$$

with τ a time-step parameter and k the iteration number. Using a finite difference approximation for the spatial derivatives, we end up with a linear system of equations involving the coefficient matrix A^{curv} which is highly structured and allows for fast solution schemes. As pointed out in [8, §12], a discrete cosine transformation can be used to diagonalize this matrix, such that the overall complexity of the solver is $\mathcal{O}(N \log N)$, with N the number of voxels. Since we solve the linear system directly and up to machine precision at each iteration, we do not need stopping rules or a convergence analysis as would be necessary for alternative iterative solution schemes, such as Multigrid, Conjugate Gradient, or SOR-like approaches.

A proper choice for the regularization parameter α depends on the order of magnitude of the forces $f^{R,T}$ and is investigated in Section 3 below. The time step τ is adjusted at each iteration such that the maximal voxel displacement $\max_x |u^{(k+1)}(x) - u^{(k)}(x)|$ is smaller than $\Delta u = 1$ voxel. Regridding and template propagation are used as in [1] to assure that the Jacobian of the deformation field is non-negative such that topology is preserved.

The method was implemented in Matlab, with image resampling, histogram computation and discrete cosine transformation coded in C.

3 Validation

We evaluated the performance of the method using the validation approach of [1] and compared it with the viscous fluid regularizer described in [1]. This approach solves

$$\nabla^2 v + \nabla(\nabla v) + f^{R,T}(x, u(x)) = 0 \quad \text{for all} \quad x \in \Omega \tag{12}$$

with $f^{R,T}$ the force field (7) and v the deformation velocity experienced by a particle at position x. An approximate solution of (12) is obtained by convolution with a Gaussian kernel ψ_γ:

$$v = \psi_\gamma \star f^{R,T} \tag{13}$$

with the spatial extent γ of the kernel controling the smoothness of the deformation. The deformation field $u^{(k+1)}$ at iteration $(k+1)$ is found by integration over time:

$$\Delta u^{(k)} = v^{(k)} - \sum_{\ell=1}^{3} v_\ell^{(k)} \left[\frac{\partial u^{(k)}}{\partial x_\ell} \right], \tag{14}$$

$$u^{(k+1)} = u^{(k)} + \Delta u^{(k)} \cdot \tau \tag{15}$$

For implementation details, we refer to [1].

Both methods were applied to simulated images generated by the BrainWeb MR simulator [2], which were non-rigidly deformed by known deformation fields u^*. These were generated by matching the T1-weighted BrainWeb image to real T1-weighted images of three periventricular leukomalacia patients, typically showing enlarged ventricles, using the viscous fluid scheme (12). The T1 and T2-weighted BrainWeb images were deformed by u^* and the original T1-weighted

image was matched to the artificially deformed images using both methods. The recovered deformation u and the ground truth u^* were compared by their root mean square (RMS) error Δu evaluated in millimeter over all brain voxels Ω_B:

$$\delta u = \sqrt{\frac{1}{\Omega_B} \sum_{x \in \Omega_B} (u(x) - u^*(x))^2} \tag{16}$$

Figure 1 illustrates the T1/T1 registration results obtained with the curvature based regularizer with $\alpha = 55$ and with the viscous fluid regularizer with $\gamma = 3$. Error values for 3 different cases are summarized in Table 1. The deformation fields obtained with $\alpha = 55$ and $\alpha = 150$ are shown in Figure 2, illustrating the smoothing effect of increasing α. Figure 3(a) plots the MI registration criterion during iterations for the curvature based regularizer. Final values of MI are 1.89 and 1.73 for $\alpha = 55$ and $\alpha = 150$ respectively. Subsequent iterations with no change in the MI criterion indicate that regridding is performed to maintain grid topology. Figure 3(b) plots the registration error obtained for the experiment in Figure 1 with the curvature based regularizer as a function of α. The error is minimal around $\alpha = 55$ and gradually increases when α is increased and more smoothing is applied.

4 Discussion

A novel free-form registration algoritm is presented that combines the MI similarity measure of [1] with the curvature based regularizer of [4] and which is implemented using a stable and fast $\mathcal{O}(N \log N)$ iterative scheme. The curvature regularizer has the advantage not to penalize affine linear transformations and was found in [4] to outperform the elastic and fluid approaches in combination with the sum of squared differences distance measure. The method was evaluated against the viscous fluid regularizer presented in [1] using artificially deformed images, with the ground truth deformations generated by the viscous fluid scheme.

As shown in Table 1, both approaches allow to recover the ground truth deformation with subvoxel RMS error in almost all experiments, both for T1/T1 as well as for T1/T2 registration. In general, smaller errors were obtained with the viscous fluid scheme which also generated the ground truth deformation. However, computing the ground truth deformation with the curvature based

Table 1. Root mean square error Δu 16 in voxels for T1/T1 and T1/T2 registrations in three different cases using the curvature based scheme C presented here and the viscous fluid scheme V of [1].

	T1/T1		T1/T2	
	C	V	C	V
Case 1	0.534	0.384	0.907	0.577
Case 2	0.612	0.304	0.769	0.443
Case 3	0.807	0.351	1.160	0.505

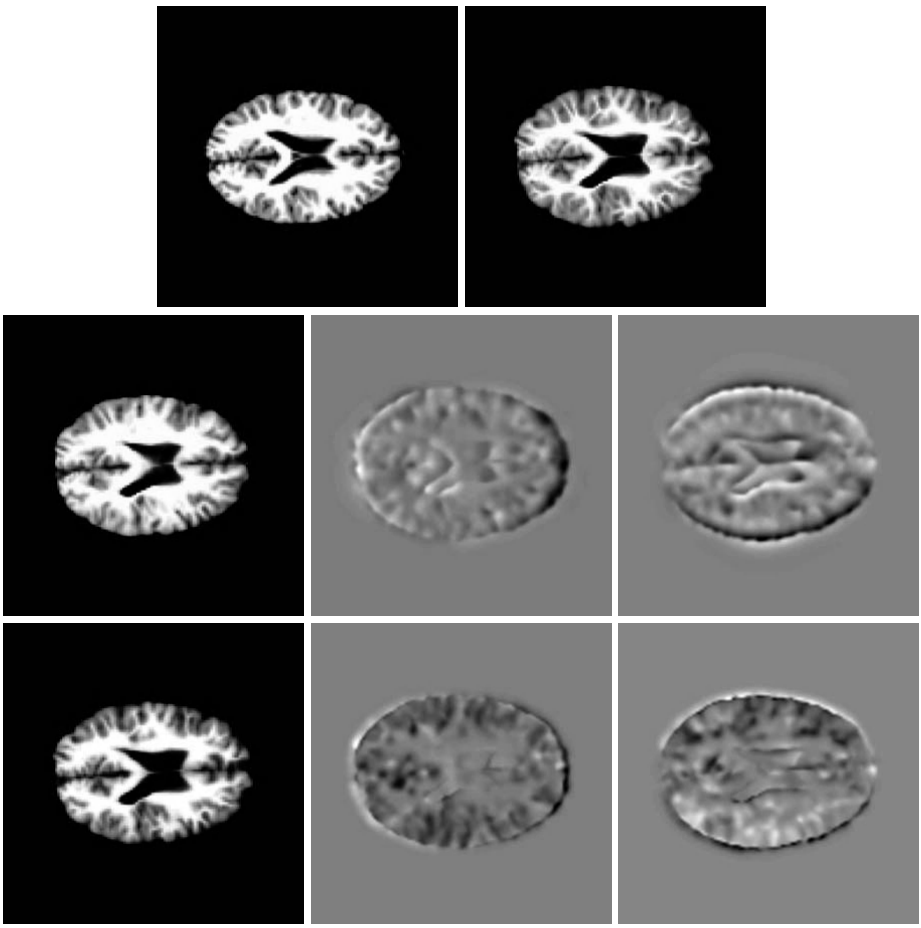
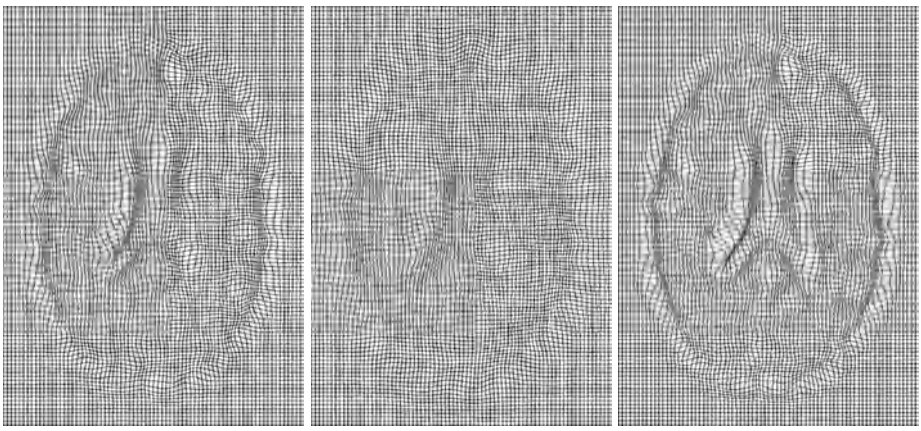


Fig. 1. Top: T1-weighted template image (left) and artificially deformed target image (right). Middle: Template matched to target using the curvature based regularizer with $\alpha = 55$ (left) and difference between ground truth and recovered deformations image (middle: horizontal component; right: vertical component). Bottom: Idem for the viscous fluid regularizer with $\gamma = 3$.

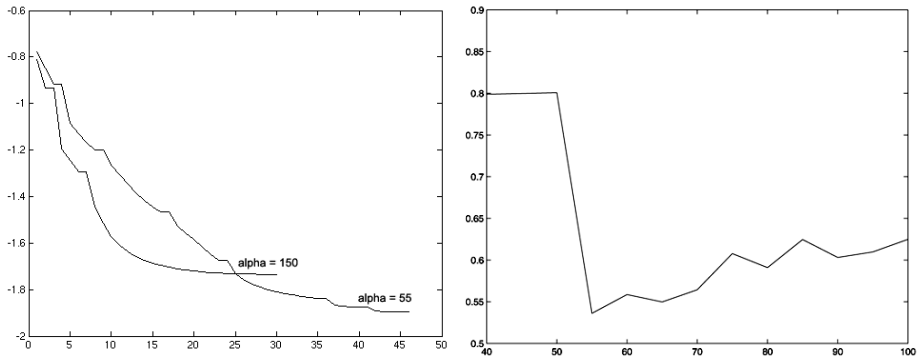
scheme for the first case T1/T1 registration yields errors of 0.539 and 1.253 voxels for the curvature and viscous fluid regularizers respectively. As illustrated in Figure 2, the curvature based regularizer yields smoother deformation fields than the viscous fluid scheme, which can be explained by the fact that the curvature constraint penalizes the second order derivatives of the deformation field. The curvature scheme can be tuned by the regularization parameter α. We studied the influence of this parameter α, determining the trade-off between smoothness of the solution and similarity with respect to the known solution.

Our results demonstrate that different regularization schemes can be applied in conjunction with the same MMI driving force field as described above, which



(a) (b) (c)

Fig. 2. Recovered deformation fields for the experiment of Figure 1 using the curvature based regularizer with $\alpha = 55$ (a) and $\alpha = 150$ (b) and using the viscous fluid regularizer with $\gamma = 3$ (c).



(a) (b)

Fig. 3. (a) Mutual information registration criterion over iterations for curvature based regularizer with $\alpha = 55$ and $\alpha = 150$; (b) Registration error (in voxels) for the experiment of Figure 1 for the curvature based regularizer with α varying between 10 and 100.

may generate equivalent registration results. Future work will focus on evaluating the ability of different schemes to generate consistent deformations that would allow to construct statistical deformation models.

References

1. E. D'Agostino, F. Maes, D. Vandermeulen, P. Suetens. A viscous fluid model for multimodal non-rigid image registration using mutual information. *MICCAI 2002*, Tokyo, 23-26 September, 2002.

2. BrainWeb MR simulator. Available at http://www.bic.mni.mcgill.ca/brainweb/.
3. B. Fischer and J. Modersitzki. A unified approach to fast image registration and a new curvature based registration technique. Preprint a-02-07, Institute of Mathematics, Medical University of Lübeck, 2002.
4. B. Fischer and J. Modersitzki. Curvature based image registration. *JMIV*, 18(1):81–85, 2003.
5. W. Hackbusch. *Partial Differential Equations*. Teubner, Stuttgart, 1987.
6. G. Hermosillo. *Variational methods for multimodal image matching*. Phd thesis, Université de Nice, France, 2002.
7. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Trans. Medical Imaging*, 16(4):187–198, 1997.
8. J. Modersitzki. *Numerical Methods for Image Registration*. Oxford University Press, to appear 2003.
9. K. Rohr. *Landmark-based Image Analysis*. Computational Imaging and Vision. Kluwer Academic Publishers, Dordrecht, 2001.
10. D. Rueckert, L.I. Sonoda, C. Hayes, D. Hill, M.O. Leach, and D.J. Hawkes. Non-rigid registration using free-form deformations: application to breast MR images. *IEEE Trans. Medical Imaging*, 18(8):712–721, 1999.

Formulation and Evaluation of Variational Curve Matching with Prior Constraints

Brian Avants and James Gee

Departments of Bioengineering and Radiology
University of Pennsylvania
Philadelphia, PA 19104-6389
{avants,gee}@grasp.cis.upenn.edu

Abstract. This work explores *a priori* constraints based on human landmarking for defining the parameters of a variational curve registration algorithm. The result is a method for designing variational energies that adjust the optimization process over the solution domain such that features which are salient in the human decision-making process are used. The application here is locating correspondence of corpora callosa that agree with expert user data. General principles that guide our particular application are first stated. These principles involve the definition of a generic variational problem and an associated optimization algorithm, given here for the case of matching curves. A small set of specific similarity criterion for curves is then defined. The ability of each feature to find correspondences that relate to human decisions is individually tested. Following the results of this study, a Maximum a Posteriori (MAP) automatic landmarking and correspondence method is developed. The probabilities associated with the MAP estimates are also used to define variational weights that vary over the solution's domain. These methods are both evaluated with respect to known correspondences and inter-user variability.

1 Introduction

Prior models are becoming a valuable asset in the design of algorithms that approximate human performance. Using prior knowledge allows one to provide a specific context in which to locate a solution. Active shape models [1,2] in particular are gaining a large amount of support for this task. These methods, however, often have unintuitive optimization criterion that may not be directly related with what drives the human registration process [3]. Furthermore, the statistical model parameters each encode global information making local control of the solution difficult. Choosing a training set that neither over nor underconstrains the model is also difficult. Similar problems arise with neural networks or other "black box" statistical methods.

An alternative to generic statistical models is given by variational energies that use prior knowledge. The advantage of this approach is that one may explicitly design variational models that take into account the human perceptual

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 21–30, 2003.

process. The difficulty is the selection of the similarity energies and then choosing their relative contributions to the solution process. Constant weighting on similarity functions are most common. However, increasing a constant weight on an individual term will affect the global solution. Locally variable weights are thus desirable. In this work, *a priori* methods for determining locally variable weights are investigated with respect to variational curve matching (VCM) [4,5,6,7].

A goal of this work is to study the properties of a group of VCM similarity functions with respect to expert anatomical correspondences. A small set of similarity functions are given for this purpose, though many others are available, for example those given in [8]. The second goal is to propose an approach for including prior information from evaluation studies to form a novel, generalizable prior model for VCM and other variational techniques. Basic principles for generating such a study as well as properties of the associated solution are proposed. These principles are used in an example with VCM. A result of their application is also given.

Overview of Some Principles for Incorporating Prior Models with Variational Optimization. It is emphasized that the following guidelines may also aid in designing variational energies for problems other than curve matching. The general approach is:

- Define an optimization problem and a method for its solution.
- Construct variational energies which measure different salient features (descriptor functions), as well as a regularization energy.
- Minimize each of the energies independently on a training data set and evaluate their performance on a sub-space in the solution domain (landmark points or regions).
- Define a statistical distribution function (s.d.f.) that approximates the likelihood of a particular sub-space with respect to :
  - the position of the sub-space in the solution domain (here, the position of landmarks across the curves' arc length),
  - the value of the similarity terms in that sub-space.

  Associate each energy with the sub-space of the solution domain on which it best performs. If no particular region exists, use regularization.
- Define *prior weights* that vary across the solution domain. When optimizing parameters that lie near a landmark (as determined by its prior), the *a priori* best energy term for that landmark is used.

This approach is intended to combine similarity functions such that the performance is optimized (as judged by consistency with a gold-standard). Further, the algorithm should preserve good solutions obtained independently, but should be more efficient than computing and combining independent solutions. Establishing prior models for similarity functions will also aid in detection of outliers. An example and validation of this approach for curve matching is now given. The remainder of the paper is organized to follow the outline above.

2 Variational Curve Matching

Curve Matching Formulation. The variational curve matching formulation, below, is most similar to that given in [6,7,4] and was previously described in [9].

Given a pair of curves, (C_1, C_2), a reparameterization is desired so the pair is better aligned with respect to some cost. The reparameterization curve is denoted $g \in \mathcal{C}_R^p \colon \xi \in \tau_g = [0,1] \rightarrow \tau_1 \times \tau_2$, where $\mathcal{C}_R^p$ denotes the p-continuous regular curves. Then, $g(\xi) = (t_1(\xi), t_2(\xi))$ which is a pairing of curve parameters for (C_1, C_2). The dependence of t on ξ is suppressed from here except where necessary. Note that a monotonicity constraint on g preserves the topology of the curves under reparameterization. That is, if a curve does not self intersect, no self intersections should be created (for example, by a reordering of the points.) Symmetry, here, is guaranteed due to the arbitrariness of labelling t_1 and t_2. An example of a reparameterization is found in figure 1.

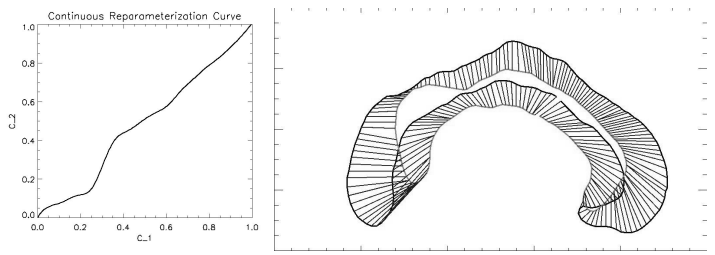


Fig. 1. A reparameterization curve, giving correspondence in the domain, is shown on the left. The axes represent the curve parameters and vary from 0 to 1. The associated correspondence in the range is shown at right.

Minimizing variational energies allows reparameterization of curve pairs with respect to geometric features. We first form the energy space, E, over which the variational problem is defined,

$$E \colon \tau_1 \times \tau_2 \rightarrow \mathbb{R}^+. \tag{1}$$

The Euclidean distance between a pair of real valued functions (f_1, f_2) defined over the arc lengths of (C_1, C_2) gives a valid energy space. Then, for any pairing (t_1, t_2) of points on the curve pair, we have some value for E, which depends on the choice of (f_1, f_2). The reparameterization curve, g, will then be a minimum energy path in E, such that,

$$g = \text{argmin } E(f_1, f_2, g). \tag{2}$$

The minimization is performed with dynamic programming. Application of the optimal g will then reparameterize the curves such that their points are aligned with respect to the features measured by E. Three examples of these types of energies are discussed below and shown in figure 2.

3 Curve Matching Energy Functions

Descriptor functions, (generally noted f_i or $f_i(t)$), for use in measuring integral similarities are now discussed.

Membrane Energy (VF). Morphometric measurements rely on having smooth mappings into the normalization space, as quantities derived from the vector field, such as the Jacobian, are the focus of analysis [10]. The smoothness of the results directly affects the statistical significance of one's study. It is therefore desirable to develop a variational energy derived from a smoothness constraint on the vector field mapping between the curve pair. Solutions to Laplace's equation,

$$\nabla^2 V(\xi) = 0, \tag{3}$$

are well known to be extremely smooth and well-behaved. We may thus choose to find a g that minimizes the equivalent integral form of Laplace's equation,

$$E_{VF}(g) = \frac{1}{2} \int_0^1 \|\nabla V(\xi)\|^2 \, d\xi, \tag{4}$$

where

$$V(\xi) = C_2 \circ t_2 - C_1 \circ t_1, \tag{5}$$

and the gradient is,

$$\nabla V(\xi) = \sum_{d=1}^{D} \frac{\partial}{\partial \xi} (C_2 \circ t_2 - C_1 \circ t_1) \hat{d}. \tag{6}$$

The $\hat{d}$ is the d^{th} dimension unit vector. Note that this is related to penalizing the change of the curves' tangents with respect to reparameterization g. This energy serves as our primary regularization term. An example of its use is in figure 1.

Distances to Geometric Structures. The variation of the minimum distance to a structure of interest can provide an important descriptor function relating to salient features. A general statement of such a distance may be denoted,

$$D(C(t), S) = \|C(t) - S_p\|, \tag{7}$$

where S_p denotes the point on geometric structure S which is most proximal to $C(t)$. Two uses of this distance follow.

An energy function that measures the distance between a curve and its center of mass may be useful for finding points that are salient to a human observer, as they are associated with the Euclidean (vs. geodesic) length of the object,

$$E_{CM}(g) = \int (D(C_1 \circ t_1, CM_1) - D(C_2 \circ t_2, CM_2))^2) d\xi. \tag{8}$$

Here, CM denotes the center of mass and is constant with respect to other parameters. Note that this function is invariant to rotation and translation and depends linearly on scale.

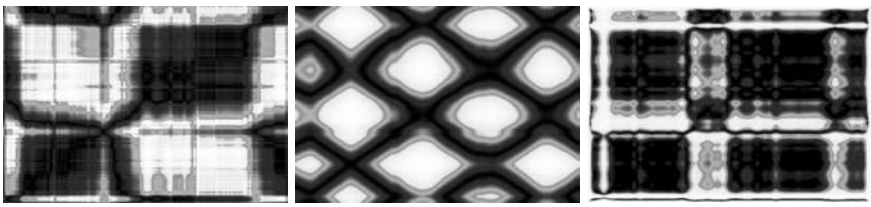


Fig. 2. Similarity functions in the energy space E of equation 1. The energy E_{VF} (left), E_{CM} (center) and E_{SS} (right). Darker color indicates lower energy.

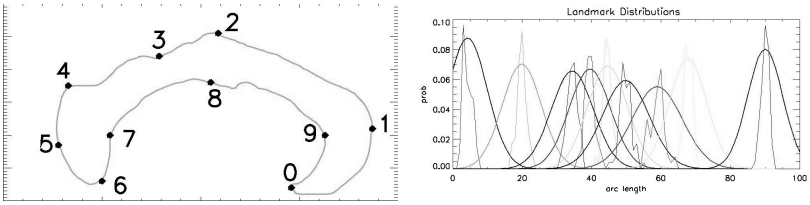


Fig. 3. Callosal atlas and landmark distributions. The human-labeled atlas used for the validation study is at left. The numbers indicate the label used to refer to each landmark. The histogram and variance-representative Gaussian distribution of the landmarks across the arc-length is shown as well.

Curvature Descriptor Function. The single-scale measure of curvature directly penalizes differences in curvature. A recently proposed scale-space curvature [11] image indicated better performance than curvature comparison, both in terms of robustness to noise and the ability to intuitively rank similar shapes [9]. This curvature scale-space image is defined as the curvature along the curve parameter versus the scale parameter. To weight each scale equally, the *curvature conservative scale-space image* preserves the integral of the curvature across scales. Denote the non-smoothed curvature function as κ and its integral over the domain of the curve as $\bar{\kappa} = \int_0^1 \kappa(t)dt$. The curvature conservative scale space function is then defined as, $\Pi(t,\sigma) = \frac{\kappa(t,\sigma)}{\int \kappa(t,\sigma)}\bar{\kappa}$, where $\kappa(t,\sigma)$ denotes the curvature of the curve at scale σ. Given this definition, Euclidean distance measures the difference of a pair (Π_1, Π_2) at parameter pair (t_1, t_2). The variational energy is,

$$E_{SS}(g) = \int (\|\Pi_1 \circ t_1 - \Pi_2 \circ t_2\|)^p d\xi. \tag{9}$$

We have found that a good choice of p for localization of landmarks is 2, which effectively makes the energy "valleys" steeper than given by the norm alone.

4 Evaluation of Variational Energies on Training Data

Landmark placement by minimization of the energies given in the previous section is now compared to human landmarking. Two expert users manually landmarked a dataset of 63 corpora callosa segmented from a set of cranial MRIs

[10]. Ten salient features on the corpus callosum were used as landmarks, as shown in figure 3. The consistency of the landmarks was analyzed between the pair of human users (Table 1, column 1) by measuring the distance between individual landmark placements. The results of VCM with each energy are also summarized in Table 1 (SD indicates standard deviation). Landmark 0 was the most consistently labelled. The VF energy provided the overall best results. The other energies were, however, more successful than VF on smaller subsets of the landmarks, as indicated in the last column of Table 1. The least consistent VF results were found for landmarks lying along regions of the callosa that exhibit less bending (landmarks one, three and four). Pure curvature similarity [4] was also evaluated, but did not outperform the energies here.

Table 1. Correspondence Method Variability (units in pixels).

Landmark	Human Mean, SD	VF Mean, SD	Best Mean, Energy
0	0.85, 0.62	0.38, 0.6	0.4, VF
1	1.53, 1.49	2.96, 2.3	2.2, CM
2	4.47, 3.59	3.60, 3.1	3.6, VF
3	3.02, 3.84	8.81, 6.9	5.0, SS
4	2.47, 2.70	4.85, 4.6	4.8, VF
5	7.22, 4.66	2.80, 2.6	2.8, VF
6	2.32, 1.73	5.32, 2.1	2.2, SS
7	1.9, 1.70	1.74, 1.7	0.9, SS
8	2.5, 2.27	3.0 , 3.7	2.8, CM
9	0.79, 0.87	1.42, 1.5	1.1, SS

One may observe that each of the three ({ VF, CM, SS }) similarity terms performed best on a different set of landmarks. The CM energy located landmarks 1 and 8 more robustly than the others as those landmarks lie on maxima and minima of the CM functions from which the CM energy is defined. Similar statements may be made about the SS energy which best performs on landmarks the exhibit high positive or negative curvature. The VF (regularization) energy performs well where the other methods do not, by minimizing the deformation energy between the pair of shapes.

5 Variational Optimization with Prior Model

The study above shows that it is possible to combine different variational optimization strategies to converge toward human decisions about correspondence. Although it is possible to run each algorithm and then select the *a priori* best algorithm's decision for each landmark, pasting the solutions together afterwards, a more efficient approach is to use prior knowledge about landmark positions (figure 3). A generic variational energy that allows natural parameter tuning is given as,

$$E(g) = \sum_i \frac{\omega_i}{\gamma_i} w_i(\xi) E_i. \tag{10}$$

The E_i may each be one of the given similarity measures. Each term is also multiplied by a normalization constant, ω_i/γ_i. These constants are used to weight the relative contributions of each term. The value γ_i normalizes the terms such that the value of ω_i explicitly states its relative contribution to the energy. Thus, if $\forall i\ \omega_i = 1.0$, each energy contributes equally to the value of the integral. Note also that the weights, w_i, are allowed to vary across the domain. The term, $w_i(\xi)$ should then weight E_i on parts of the domain where it is known, from an evaluation study as above, to perform well. Methods for defining functions w_i and γ_i are the subject of the following sections.

Normalizing Similarity Term Contributions. The motivation is to enable a user or algorithm designer to choose weights that correspond with numerical intuition. The condition that is desired is an equal numerical contribution to the integral energy from each term in the energy function. Toward this end, a numerical algorithm is used. Denote a permissible solution to the variational problem as g and the identity solution as g_I. The normalization method performs the following steps, initializing with $g = g_I$:

0. Given g, measure the set of integral energies $\{E_i\}$.
1. Set $\gamma_i = E_i(g)$.
2. Find a new g that optimizes Equation 10 given weights $\{\gamma_i, \omega_i\}$.
3. Go to step 0.

Note that this simple approach may be incorporated into most iterative optimization algorithms. However, this does not solve the real problem at hand, which is the fact that simply increasing weight ω_i without setting variable weights $w_i(\xi)$ has global effects on one's solution. This was illustrated with an evaluation study in which this normalization algorithm was used with the following optimization problem,

$$E(g) = c_{VF}E_{VF} + c_{CM}E_{CM} + c_{SS}E_{SS}. \tag{11}$$

Here, the constants were arrived at by a single iteration of the normalization algorithm above. It was observed by using evaluation that landmark 6 was not well-aligned by this choice of weights. Thus, due to evaluation results on independent energy functions (see table 1), it was reasoned that c_{SS} should be increased. This improved the registration of landmark 6 but resulted in compromised performance on landmarks 1 and 8.

A solution to this problem is to define the functions $w(\xi)$ in such a way as to enable localized adjustment of the similarity contributions. A function that may be used for curve matching in conjunction with the similarity energies shown in figure 2 is given in figure 4. A non-parametric estimation technique for these statistical distributions is discussed below.

MAP Distribution Estimation. The distributions of the arc-length, SS and CM descriptors with respect to the individual landmarks are now estimated. First, we collect this set of descriptor functions associated with each landmark.

That is, for each landmark on each curve, collect the curvature, center of mass distance, and arc length position associated with it. Next, use Parzen windowing [2] with that collection of training data in order to estimate a maximum *a posteriori* (MAP) probability,

$$\Pr(LM|f_i) \propto \Pr(f_i|LM)\Pr(LM). \tag{12}$$

The functions f_i denote the associated descriptor functions, f_t, f_{SS}, f_{CM}, with t the arc-length. To get the total probability of a given landmark, LM, we take the product of the MAP probabilities for each f_i. Note that this MAP probability is a function defined along the given curve. The probability of a landmark, $\Pr(LM)$, is constant. The likelihoods are individually estimated with Parzen windowing, using a Gaussian smoothing function and with the variance estimated from the data as the window width. A related method for segmentation is given in [2]. Note, that this also provides a MAP matching which is given by taking, for each landmark, the position on the curve that has maximal MAP probability. This method is tested in Table 2. Figure 4 shows the Parzen window joint probability estimates for the SS and CM descriptors. The same figure also shows the joint MAP probability used for matching a pair of curves. The joint landmark probability is gained by taking the product of the MAP probabilities for each curve, such that $\Pr_{LM}(\xi) = \Pi_i \Pr_i(LM|f_i)(t_1)\Pi_j \Pr_j(LM|f_j)(t_2)$. Here, the training was performed on the data given by User 1. The performance was then tested with respect to User 2.

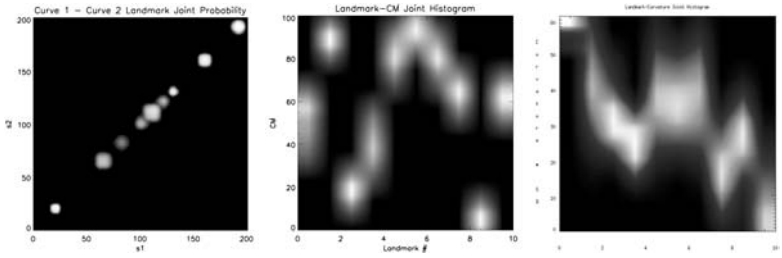


Fig. 4. Variable weight function. A weight function which varies over the curve parameter pairings is shown. In fact, this weight function is the joint landmark probability between the two curves. It is related to the probability of the CM descriptor-Landmark joint probability (center) and SS descriptor-Landmark joint probability (right).

***Prior Weight* Construction.** Construction of the *prior weights* consists of first normalizing the energy functions such that they have approximately equivalent contributions, as stated above, and then providing desired values for the ω_i. The second step is to combine the landmark MAP distribution functions with the knowledge of the performance of each energy term on each landmark to create the final *prior weight* functions, $\{w_{VF}(\xi), w_{CM}(\xi), w_{SS}(\xi)\}$. One way to use these functions is given by the integral,

$$E = (1.0 - \underset{LM6}{\Pr}(\xi))E_{VF} + (1.0 - \underset{LM6}{\Pr}(\xi))E_{CM} + \underset{LM6}{\Pr}(\xi)E_{SS}. \tag{13}$$

Table 2. MAP Algorithm-User 2 Variability.

Landmark	Mean Distance	SD	Median	Max
0	0.73	0.33	0.72	1.9
1	2.97	2.39	2.16	8.39
2	2.90	2.67	2.73	10.90
3	7.44	5.98	6.21	22.43
4	4.60	3.94	4.27	15.19
5	3.69	2.98	4.01	12.04
6	2.18	1.83	2.10	6.12
7	2.23	1.96	2.14	10.11
8	2.87	2.59	2.17	10.71
9	1.05	1.32	0	4.63

Table 3. Prior SS Algorithm-User 2 Variability.

Landmark	Mean Distance	SD	Median	Max
0	0.38	0.60	0.00	1.7
1	1.62	1.76	1.83	6.5
2	5.42	5.23	3.91	18.1
3	4.80	5.50	2.21	21.7
4	5.44	7.45	2.00	27.2
5	4.59	4.56	3.66	17.3
6	2.76	2.10	3.36	8.3
7	1.16	1.28	1.75	4.0
8	2.86	2.46	2.14	11.3
9	1.69	1.38	1.78	4.0

The function $\Pr_{LM6}$ is scaled such that it has maximum 1.0 and denotes the joint two-curve MAP probability for landmark 6 with respect to its descriptor functions. This defines an optimization criterion that attempts to allow the SS energy to guide the optimization in that region of the curve. A related criterion that uses landmarks 6, 7 and 9 prior weights was evaluated with respect to human performance. These prior weights (figure 4) were chosen to improve the performance on landmarks 6, 7 and 9 in accord with the good results of the SS energy on those parts of the domain. The results of this SS prior algorithm (below in table 3) show very similar median performance with the human users. The means, however, are not as similar due to failure in some cases (large maximum error) for landmarks 3 and 4. The results of this prior-weighted variational energy outperformed any of the single term energies, regardless of varying the relative constant weightings.

6 Discussion and Conclusion

Generic principles for using performance evaluation to define prior-constrained variational methods were given. The paper showed results of variational energy

evaluation for membrane, scale-space curvature and center of mass distance energy functions. It was shown that each energy performed best on some subset of the landmark set. The goal of integrating these competing optimization terms in a single optimization condition was then investigated. A non-parametric statistical model using Parzen windowing located MAP distributions for the landmarks. These distributions were then leveraged to provide locally variable weights yielding a locally adaptive variational solution. This three-term energy integral out-performed any single term algorithm that was evaluated and also closely approximated inter-human variability. Future work will focus on furthering rigorous Bayesian formulations and distribution estimation techniques to create better adaptive weight models for variational optimization of segmentation and registration methods.

References

1. T.F. Cootes, C.J. Taylor, and J. Graham, "Active shape models - their training and application," *Computer Vision and Image Understanding*, vol. 60, pp. 38–59, 1995.
2. M. Leventon, E. Grimson, and O. Faugeras, "Statistical shape influence in geodesic active contours," in *Proc. IEEE Conf. Comp. Vision and Patt. Recog.*, 2000, pp. 4–11.
3. R. Davies, T. Cootes, J. Waterton, and C.J. Taylor, "An efficient method for constructing optimal statistical shape models," in *Medical Image Computing and Computer Assisted Intervention*, 2001, pp. 57–65.
4. T. Sebastian, P. Klein, B. Kimia, and J. Crisco, "Constructing 2D curve atlases," in *Mathematical Methods in Biomedical Image Analysis*, 2000, pp. 70–77.
5. L. Younes, "Computable elastic distance between shapes," *SIAM J. Appl. Math*, vol. 58, pp. 565–586, 1998.
6. H.D. Tagare, D. O'Shea, and A. Rangarajan, "A geometric criterion for shape based non-rigid correspondence," in *Fifth Intl. Conf. on Computer Vision*, 1995, pp. 434–439.
7. D. Geiger, T. Liu, and R. Kohn, "Representation and self-similarity of shapes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, pp. 86–99, 2003.
8. T. Liu and D. Geiger, "Approximate tree matching and shape similarity," *International Conference on Computer Vision*, pp. 456–462, 1999.
9. B. Avants and J.C. Gee, "Continuous curve matching with scale-space curvature and extrema-based scale selection," in *Scale-Space Theories in Computer Vision*, 2003, p. in press, L. Griffin editor, Heidelberg:Springer-Verlag, LNCS.
10. A. Dubb, B. Avants, R. Gur, and J. C. Gee, "Shape characterization of the corpus callosum in Schizophrenia using template deformation," in *Medical Image Computing and Computer-Assisted Intervention*, R. Kikinis, Ed., Heidelberg, 2002, pp. 381–388, Springer-Verlag.
11. F. Mokhtarian and A. Mackworth, "Scale-based description and recognition of planar curves and two-dimensional shapes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, pp. 34–44, 1986.

Normalized Mutual Information Based PET-MR Registration Using K-Means Clustering and Shading Correction

Z.F. Knops[1], J.B. Antoine Maintz[1], M.A. Viergever[1,2], and J.P.W. Pluim[2]

[1] Utrecht University, Department of Computer Science, PO Box 80089,
NL-3508 TB, Utrecht, The Netherlands
{zeger,twan}@cs.uu.nl
[2] University Medical Center Utrecht, Image Sciences Institute, E01.335, PO Box 85500,
Utrecht NL-3508 GA, Utrecht, The Netherlands
{max,josien}@isi.uu.nl

Abstract. A method for the efficient re-binning and shading based correction of intensity distributions of the images prior to normalized mutual information based registration is presented. Our intensity distribution re-binning method is based on the K-means clustering algorithm as opposed to the generally used equidistant binning method. K-means clustering is a binning method with a variable size for each bin which is adjusted to achieve a natural clustering. Furthermore, a shading correction method is applied to reduce the effect of intensity inhomogeneities in MR images. Registering clinical shading corrected MR images to PET images using our method shows that a significant reduction in computational time without loss of accuracy as compared to the standard equidistant binning based registration is possible.

1 Introduction

Several different imaging modalities are used in current clinical practice. Different modalities can provide information about the physical properties of anatomical structures and tissues or provide functional information about the use of those structures and tissues. It is clear that the combination of these modalities can provide us with extra information. In order to combine the information of different modalities we have to align those images spatially. Spatial alignment, registration, is important in a number of applications for the diagnosis, therapy and monitoring of disease and for research on specific anatomical structures.

In order to do automated registration a quality measure of registration is necessary in order to find the correct spatial alignment of the images. In this paper we use normalized mutual information (NMI) [1][2][3][4][5] and rigid transformation. NMI has proven to be a robust an accurate similarity measure for many types of mono- and multi-modality image registration.

In some MR images slow intensity variations of the same anatomical tissue over the image domain occur. These intensity inhomogeneities have adverse effects when registering those images, both the computational time and the probability of misregistrations can increase. Many solutions to counter inhomogeneities have been proposed. In this paper we use the method as proposed by Likar et al. [6][7] which is based on entropy minimization. Likar describes the image degradation by a linear model con-

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 31–39, 2003.

sisting of a multiplicative and an additive component in combination with smoothly varying basis functions. These parameters are optimized by reducing the entropy while preserving the global intensity statistic. A method that improves entropy is especially desired in this context, because the NMI registration measure is also entropy-based.

We would like to find the optimal registration of two images which is defined as the spatial transformation that yields the highest NMI value. During the registration process we need an estimation of the intensity distributions of the images. Equidistant binning is most commonly used, where a priori a fixed number of so-called bins are chosen, and each intensity is mapped in a rounded linear fashion to the bins of the joint histogram. When using equidistant binning, natural image segments could end up in different bins. This splitting of segments possibly has adverse effects on the registration procedure in terms of computation time and probability of misregistration. We use a different binning method, K-means clustering [8], which uses a variable bin size for each bin in order to achieve a natural clustering. The results of the intensity inhomogeneity correction are likely to enhance the K-means clustering method due to the reduction of the histogram dispersion [9].

In section 2 we will discuss the registration process with NMI and introduce our approach with respect to the binning method and inhomogeneity correction. In the section 3 the materials are presented. In section 4 we will discuss the experimental results of registration with our new binning approach. Our conclusions are presented in section 5.

2 Methods

2.1 Normalized Mutual Information

For a discrete random variable A, where $p_A(a)$ is the probability that A has value a, the Shannon entropy H is defined as

$$H(A) = -\sum_a p_A(a) \log p_A(a). \tag{1}$$

The entropy for an image is a measure of how well we can predict the intensity at an arbitrary point in the image. If the entropy for an image is zero we have no uncertainty about the intensity at an arbitrary point in the image and it is clear the image is completely homogeneous. On the other hand if the entropy for an image is high we can assume the image consists of a large number of equally likely intensities. Note that the entropy does not take any spatial information into account.

For two discrete random variables A and B we can define the Shannon entropy of their joint distribution as

$$H(A,B) = -\sum_{a,b} p_{AB}(a,b) \log p_{AB}(a,b). \tag{2}$$

To compute the Shannon entropy of the registration of two images the joint-histogram can be used. A joint-histogram is an estimation of the distribution of the intensity couples on corresponding voxels in the images.

In this study we used normalized mutual information, NMI, which is a well established registration quality measure and can be computed using the image entropies [10][11]:

$$NMI(A,B) = \frac{H(A) + H(B)}{H(A,B)}. \tag{3}$$

2.2 Intensity Inhomogeneity Correction in MR Images

In some MR images slow intensity variations of the same tissue class over the image domain occur. There can also be spurious intensity variations which may reach up to 30% of the image intensity amplitude. These intensity inhomogeneities can be caused by poor radio frequency (RF) coil uniformity, static field inhomogeneity, RF penetration, gradient-driven eddy currents, and overall patient anatomy and position [6].

Likar [6] uses a retrospective shading correction based on entropy minimization. The algorithm is based on the assumption that an image with inhomogeneities, *P*, has higher entropy, contains more information, than the same image, *A*, without inhomogeneities.

$$H(P) = H(f(A)) > H(A) \tag{4}$$

We try to find a correction f^{-1} which is constrained not to change the mean intensity of *A* and can not transform the input image to a uniform image. A multiplicative and additive model is used to describe the inhomogeneities.

$$P = f(A) = f_{Mul}A + f_{Add} \tag{5}$$

Both f_{Mul} and f_{Add} are modeled as five-parametric quadratic functions. The optimal parameters are found using Powell's method [12].

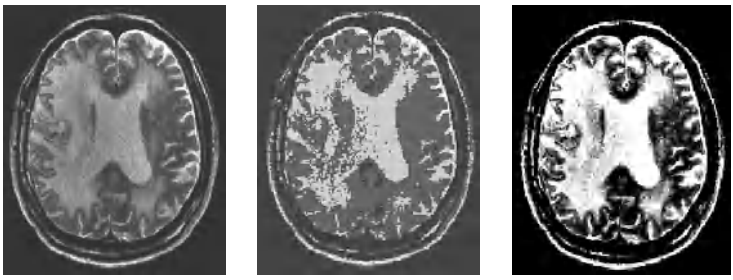
2.3 Non-equidistant Binning by Clustering

For both equidistant binning and K-means clustering the image histogram is used. Equidistant binning does not take the entries of the histogram into account in contrast to K-means clustering, where the length of each interval is adjusted according to the variance of the intensities in that interval.

With equidistant binning of an intensity distribution it is possible anatomical structures end up in different bins. A peak representing a distinct anatomical structure could be split into two bins. By adjusting the size of each bin we can achieve a natural clustering and less dispersion in the joint-histogram. The optimal bin size is found by minimizing the variance of intensities within each bin.

2.4 Clustering with Dynamic Programming

We use dynamic programming [13] to find the optimal clustering for an image. The solution for K-means clustering can be divided into separate smaller subproblems which take less computational time to solve.



(a) (b) (c)

Fig. 1. Part of a slice of a 3D MR dataset (a). Effects of equidistant binning with 8 bins are visible in image (b) segments that are clearly visible in (a) have been removed or altered. Image (c) is the result of the K-means clustering binning method. Segments that are clearly visible in (a) are not altered or removed. Contrast has been adjusted in all images to emphasize the differences.

The results of an equidistantly binned and a K-means clustered binned MR image using 8 bins are shown in figure 1. Contrast has been adjusted in all 3 images to emphasize the differences. Segments that are clearly visible in the original image (a) have been removed or altered in the equidistantly binned image (b). The K-means clustered image shows a clear correspondence with image (a), segments have not been altered or removed.

Using intensity inhomogeneity correction based on lowering the entropy of the image, such as in Likar's method, results in a more sharply peaked, less dispersed histogram. Less dispersion in the histogram enhances the K-means clustering based binning. The K-means clustering algorithm is more likely to find anatomical structures as segments in the histogram and thus is more likely to enhance the registration process.

3 Materials

One set of data comprising seven 3D PET/MR patient studies has been used in this study: the RREP data [14]. The dataset comes with a reliable gold standard for registration. Furthermore, the RREP dataset is an internationally accepted standard dataset for comparing rigid registration methods.

RREP data (7 patient studies)
MR T2 weighted data consists of 4.0mm thick slices with a pixel size of 1.25mm square. All the slices have a resolution of 256 pixels square and the number of slices varied between 20 and 26. The echo time was 90ms and the repetition time 3000ms.

PET data consists of 8.0mm thick slices with a pixel size of approximately 2.59mm square. All the slices have a resolution of 128 pixels square, the number of slices was 14. F-18 was used as a radionuclide.

In the following median and average results of the experiments performed on all seven studies are presented.

4 Results and Experiments

Equidistant binning and K-means clustering were compared with and without inhomogeneity correction for rigid registration of 3D images. To optimize the NMI measure as a function of rigid 3D geometric transformations applied to one of the images Powell's method [12] was used. This method optimizes the NMI by using one-dimensional optimization steps. The number of optimization steps per registration can be used as a measure of computational speed. We measured the number of steps needed to converge per registration for both binning and K-means clustering, both with and without inhomogeneity correction. The quality of the solution found was assessed using a known gold standard. Registrations were done for a variety of numbers of bins.

To compare the solution found with the gold standard, a sphere was fitted approximately around the head with 10.000 points on the boundary. For each point the distance between the original gold standard and the position transformed with the transformation of our solution was taken. The median of these distances was taken as the error measure. If the median error was above 8mm, the slice thickness of the PET datasets, the registration was considered to be a misregistration.

4.1 Optimization Steps

For all experiments the identity matrix was used as initialization of Powell's method with a step size of 0.5 for each transformation parameter. In the case of misregistration no attempts were made to find a correct registration by adjusting the step size.

The number of optimization steps needed to reach registration for PET-MR without inhomogeneity correction are compared for equidistant binning and K-means clustering in table 1. The reduction in optimization steps is given as a percentage and is achieved by using K-means clustering (C) based binning instead of equidistant binning (B).

Table 1. Comparison of the number of optimization steps, median error and the number of misregistrations of registration using K-means clustering (C) versus registration using binning (B). The errors are given in mm and the reduction in optimization steps is in %, both are computed for the non-misregistrations only.

C vs. B	Number of bins					
	4	**8**	**16**	**32**	**64**	**128**
Reduction in optimization steps	-9	12	92	21	55	-3
Average median error C	3.76	3.39	4.50	3.67	5.61	3.32
Average median error B	4.67	3.10	3.73	3.35	2.49	3.18
Average median error difference C-B	-1.37	0.51	0.23	0.14	1.89	0.40
Number of misregistrations C	0	0	1	1	4	4
Number of misregistrations B	3	2	3	3	4	3

For 8, 16, 32 and 64 bins K-means clustering requires on average fewer optimization steps than equidistant binning and the average median error differences are

within 2mm of each other. Using 16 or 32 bins yields the best results; there is a reduction in the number of misregistrations, a substantial decrease in optimization steps and the average median error differences are within 0.23mm. Overall, there were 10 misregistrations for clustering and 18 for binning.

The number of optimization steps needed to reach registration for PET-MR are compared for equidistant binning and K-means clustering both with inhomogeneity correction in table 2. The percentages indicate the reduction in optimization steps achieved by using K-means clustering with inhomogeneity correction (IC) based binning instead of fixed binning with inhomogeneity correction (IB).

Table 2. Comparison of the number of optimization steps, median error and the number of misregistrations of registration using K-means clustering with inhomogeneity correction (IC) versus registration using binning with inhomogeneity correction (IB). The errors are given in mm and the reduction in optimization steps is in %, both are computed for the non-misregistrations only.

IC vs. IB	Number of bins					
	4	**8**	**16**	**32**	**64**	**128**
Reduction in optimization steps	23	-28	31	129	22	1
Average median error IC	3.82	3.26	2.99	3.60	3.69	3.35
Average median error IB	4.02	3.42	4.23	3.81	3.30	4.37
Average median error difference IC-IB	-0.36	-0.07	-1.05	-0.25	0.48	-1.10
Number of misregistrations IC	0	0	1	1	0	2
Number of misregistrations IB	1	1	2	3	2	3

For 4, 16, 32, 64 and 128 bins K-means clustering requires on average fewer optimization steps than fixed binning and the average median error differences are within 1.05mm of each other. Using 4, 16, 32 or 64 bins yields the best results: there is a reduction in the number of misregistrations, a significant reduction in optimization steps and the average median error differences are within 1.05mm of each other. Overall, there were 4 misregistrations for clustering and 12 for binning.

The number of optimization steps needed to reach registration for PET-MR are compared for K-means clustering with and without inhomogeneity correction as shown in table 3. The percentages indicate the reduction in optimization steps achieved by using K-means clustering with inhomogeneity correction (IC) based binning instead of K-means clustering without inhomogeneity correction (C).

Except when using 32 bins, K-means clustering increases the number of optimizations steps needed. There is no significant difference in error for any number of bins. Only when using many bins (64 or 128) does the inhomogeneity correction reduce the number of misregistrations. There were 4 misregistrations for K-means clustering with inhomogeneity correction and 10 for K-means clustering without inhomogeneity correction.

The number of optimization steps needed to reach registration for PET-MR are compared for equidistant binning with and without inhomogeneity correction as shown in table 4. The percentages indicate the reduction in optimization steps achieved by using equidistant binning with inhomogeneity correction (IB) based binning instead of equidistant binning without inhomogeneity correction (B).

Table 3. Comparison of the number of optimization steps, median error and the number of misregistrations of registration using K-means clustering with inhomogeneity correction (IC) versus registration using K-means clustering without inhomogeneity correction (C). The errors are given in mm and the reduction in optimization steps is in %, both are computed for the non-misregistrations only.

IC vs. C	Number of bins					
	4	**8**	**16**	**32**	**64**	**128**
Reduction in optimization steps	-17	-43	-93	22	-18	-1
Average median error IC	3.82	3.26	2.99	3.60	3.69	3.35
Average median error C	3.76	3.39	4.50	3.67	5.61	3.32
Average median error difference IC-C	0.06	-0.12	-1.46	-0.08	-1.90	0.34
Number of misregistrations IC	0	0	1	1	0	2
Number of misregistrations C	0	0	1	1	4	4

Table 4. Comparison of the number of optimization steps, median error and the number of misregistrations of registration using binning with inhomogeneity correction (IB) versus registration using binning without inhomogeneity correction (B). The errors are given in mm and the reduction in optimization steps is in %, both are computed for the non-misregistrations only.

IB vs. B	Number of bins					
	4	**8**	**16**	**32**	**64**	**128**
Reduction in optimization steps	-52	7	-41	-22	19	10
Average median error IB	4.02	3.42	4.23	3.81	3.30	4.37
Average median error B	4.67	3.10	3.73	3.35	2.49	3.18
Average median error difference IB-B	-1.15	0.13	0.50	1.33	0.20	0.60
Number of misregistrations IB	1	1	2	3	2	3
Number of misregistrations B	3	2	3	3	4	3

For all number of bins, the error difference is not significant; and there is no discernable pattern in the variation of the optimization steps. However the number of misregistrations was reduced for all bin sizes except 32 and 128. Overall, there were 12 misregistrations for equidistant binning with inhomogeneity correction and 18 for equidistant binning without inhomogeneity correction.

The number of optimization steps needed to reach registration for PET-MR are compared for K-means clustering with inhomogeneity correction and binning as shown in table 5. The percentages indicate the reduction in optimization steps achieved by using K-means clustering with inhomogeneity correction (IC) based binning instead of equidistant binning without inhomogeneity correction (B).

For 4, 32, 64, 128 bins K-means clustering with inhomogeneity correction requires on average fewer optimization steps than equidistant binning without inhomogeneity correction. For all number of bins there was a reduction of the number of misregistrations, while there was no significant difference in error. Overall, there were 4 misregistrations for K-means clustering with inhomogeneity correction and 18 for equidistant binning without inhomogeneity correction.

Table 5. Comparison of the number of optimization steps, median error and the number of misregistrations of registration using K-means clustering with inhomogeneity correction (IC) versus registration using binning without inhomogeneity correction (B). The errors are given in mm and the reduction in optimization steps is in %, both are computed for the non-misregistrations only.

IC vs. B	Number of bins					
	4	**8**	**16**	**32**	**64**	**128**
Reduction in optimization steps	13	-20	-10	89	24	36
Average median error IC	3.82	3.26	2.99	3.60	3.69	3.35
Average median error B	4.67	3.10	3.73	3.35	2.49	3.18
Average median error difference IC-B	-1.34	0.31	-0.55	0.27	1.11	0.74
Number of misregistrations IC	0	0	1	1	0	2
Number of misregistrations B	3	2	3	3	4	3

5 Conclusion

In summary, the use of inhomogeneity correction significantly improves the registration process for both equidistant binning and K-means clustering based registration in terms of the number of misregistrations occurring, without loss of accuracy. The use of K-means clustering generally decreases the number of optimization steps needed, also without loss of accuracy.

The combined use of K-means clustering based binning and inhomogeneity correction yields the largest decrease in the number of misregistrations for any number of bins, again without loss of accuracy. When using 32 or 64 bins, the use of K-means clustering with or without inhomogeneity correction additionally yields a significant decrease in number of optimization steps needed.

References

1. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens and G. Marchal, “Automated multimodality image registration based on information theory”, in Information Processing in Medical Imaging, Y. Bizais, C. Barillot and R. Di Paolo (eds.), Kluwer Academic Publishers, Dordrecht (1995) 263-274
2. Collignon, Multi-modality medical image registration by maximization of mutual information, Ph.D. thesis, Catholic Universtity of Leuven, Leuven, Belgium (1998)
3. P. Viola and W. M. Wells III, “Alignment by maximization of mutual information”, in International Conference on Computer Vision, E. Grimson, S. Shafer, A. Blake and K. Sugihara (eds.), IEEE Computer Society Press, Los Alamitos CA, USA (1995), 16-23
4. P. Viola, Alignment by maximization of mutual information, Ph.D. thesis, Massachusetts Institute of Technology, Boston, MA, USA (1995)
5. W. M. Wells III, P. Viola and R. Kikinis, “Multi-modal volume registration by maximization of mutual information”, in Medical Robotics and Computer Assisted Surgery, John Wiley & Sons, New York (1995) 55-62
6. Likar, M. A. Viergever and F. Pernus. “Retrospective correction of MR intensity inhomogeneity by information minimization”. In Proc. MICCAI 2000. Springer-Verlag, Berlin Heidelberg New York (2000)

7. Likar, J. B. A. Maintz, M. A. Viergever and F. Pernus. “Retrospective Shading Correction Based On Entropy Minimization”, Vol. 197. Journal of Microscopy (2000) 285-295
8. J. MacQueen. Some methods for classification and analysis of multivariate observations, In proc. of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Volume I, Statistics. M. Le Cam and Jerzy Neyman (eds.). University of California Press (1967)
9. Z. F. Knops. J. B. A. Maintz, M. A. Viergever and J. P. W. Pluim, “Normalized mutual information based registration using K-means clustering based histogram binning”. In proc. SPIE Medical Imaging 2003, (2003) in press
10. Studholme, Measures of 3D medical image alignment, Ph.D. thesis, University of London, London, UK (1997)
11. Studholme, D. L. G. Hill and D. J. Hawkes, “An overlap invariant entropy measure of 3D medical image alignment”, Pattern Recognition, Vol. 32, no. 1. (1999) 71-86
12. W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, Numerical Recipes in C, Cambridge University Press, Cambridge UK (1992)
13. Richard Bellman, “Dynamic Programming Treatment of the Travelling Salesman Problem”, JACM, Vol. 9, no. 1. (1962) 61-63
14. http://www.vuse.vanderbilt.edu/~image/registration

Computational Anatomy and Implicit Object Representation: A Level Set Approach

Wei-Hsun Liao[1], Luminita Vese[2], Sung-Cheng Huang[1],
Marvin Bergsneider[3], and Stanley Osher[2]

[1] Department of Biomathematics, UCLA David Geffen School of Medicine, LA, CA, 90095
[2] Department of Mathematics, UCLA, LA, CA, 90095
[3] Brain Injury Center, UCLA, LA, CA, 90095

Abstract. In this paper, we look at the fundamental problem of object matching in computational anatomy. We present a new framework for warping pairs of overlapping and non-overlapping shapes, open curves, and landmarks based on the level set approach. When implemented in 3-D, the same framework could be used to warp 3-D objects with minimal modification. Our approach is to use the level set functions to represent the objects to be matched. Using this representation, the problem becomes an energy minimization problem. Cost functions for warping overlapping, non-overlapping, open curves, and landmarks are proposed. Euler-Lagrange equations are applied and gradient descent is used to solve the corresponding partial differential equations. Moreover, a general framework for linking the level set approach and the infinite dimensional group actions is discussed.

1 Introduction

Computational anatomy [1, 2] is an emerging new discipline that deals with analyzing and making sense of the large collection and database of brain imaging. A fundamental problem in computational anatomy is image warping, or dynamically mapping one brain dataset to another through diffeomorphic transformation. In this paper, we will focus on developing techniques for matching anatomically important objects.

Object warping is a challenging problem not only in computational anatomy but also in computer vision, pattern recognition as well as many other scientific fields. In the past decade, several strategies of non-rigid warping algorithms have been proposed that could be divided into two groups: landmark based and dense matching.

Landmark matching involves first identifying user-defined landmarks that need to be matched. By interpolating the discrete matching of the landmarks, one tries to obtain a dense diffeomorphism for the whole image. Dense matching starts by forming a cost function that is minimized when the objects are matched. In order to ensure smooth matching, a regularizing term on the deformation field is added.

In this paper, we will use the terms template and study to denote the images to be matched. Let us denote the template image as $T(x)$ and the study image as $S(x)$ which

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 40–49, 2003.

are images on the spatial domain $\Omega \subset R^n$. The problem of image warping is to find a displacement field $u(x)$ at each point x such that a properly defined cost function, which will be denoted by $D(T,S,u)$, between the deformed template and the study is minimized. The displacement field is a vector field such that given any displacement field u the deformed template is given by $T(x-u)$. The term displacement is used because it can be viewed as how a point in the template is moved away from its original location. The most common way to define the measure between the deformed template and the study image is based on the L^2 norm

$$D(T,S,u) = \frac{1}{2}\int_{\Omega} |\, T(x-u) - S(x) \,|^2 \, dx \,. \tag{1}$$

Gradient descent of the corresponding Euler-Lagrange equation is often used to minimize this cost function

$$\frac{\partial u(x,t)}{\partial t} = f(x,u(x,t)), f(x,u) = [T(x-u) - S(x)] \bullet \nabla T \,|_{x-u} \,. \tag{2}$$

The function $f(x,u)$, which is often called the force field or the body force, describes the derivative of the cost function with respect to the displacement field u.

2 Previous Work

Several models for regularizing the deformation field have been proposed. We will give an overview by looking at those with the most theoretical interests

2.1 Hyper-elastic Matching

In hyper-elastic matching [3], the authors tried to draw analogy between image warping and deforming elastic plates. Under the assumption of linear elasticity, which holds for relatively small displacement field only, we arrive at the following equation that should hold at equilibrium

$$\mu \Delta u(x,t) + (\mu + \lambda)\nabla(\nabla \cdot u(x,t)) = f(x,u(x,t)) \,. \tag{3}$$

Here μ and λ are the *Lame* constants. Due to this linear elasticity assumption, large-magnitude displacements are severely penalized and thus hyper-elastic model is not suitable for problems in which large and highly nonlinear deformation is needed.

2.2 The Fluid Dynamic Approach

An extension of the hyper-elastic matching, the viscous fluid matching, was proposed by Christensen [4-6] using the Navier-Stokes formulation. This approach can be viewed as the linear elastic operator applied to the velocity field, which is related to the displacement field in a standard Eulerian framework. The advantage of this method is that it allows large-magnitude deformations since stress constraining the deformation relaxes over time.

2.3 Large Deformation through Diffeomorphisms by Infinite Dimensional Group Actions

In the past decade, many researchers have tried to establish rigorous theories based on continuum mechanics that ensure diffeomorphic transformation by working on the forward and inverse mapping directly (see [7] and the references therein).

In summary, Let $g^{-1}(x)=x-u$ be the deformation field, and G be the group which is formed by all the diffeomorphisms that map $\Omega \in X$ to itself. A path in G

$$g(x,t) = g_t(x),\ g_t \in G \text{ for } \forall t \in [0,1]\ . \tag{4}$$

is linked to the velocity field v_t by the following equation

$$\frac{\partial}{\partial t} g_t(x) = v_t(g_t(x))\ . \tag{5}$$

Once the forward path g_t is defined, the inverse path is uniquely determined by

$$g_t^{-1}(g_t(x)) = x\ , \tag{6}$$

and is linked to the velocity by

$$\frac{\partial}{\partial t} g_t^{-1}(x) = -Dg_t^{-1}(x)v_t(x)\ . \tag{7}$$

Given a path $g(x, t)$ and its associated velocity, we define the energy of the path by

$$E(g) = \int_{t=0}^{1} \left\| v_t(x) \right\|_L^2 dt = \int_{t=0}^{1} \left\langle Lv_t(x), Lv_t(x) \right\rangle dt\ . \tag{8}$$

Here L is a differential operator. Furthermore, let us define the momentum p by

$$p(x,t) = p_t(x) = L^{\dagger} Lv_t(x)\ . \tag{9}$$

Here $L^{\dagger}$ is the adjoint operator of L. It can be shown that diffeomorphisms could be ensured under mild restrictions on the operator L (see [8]). We then have the following theorem which could be found in [7, 9].

Theorem 1

Let g_0 and g_1 be elements in G, the function d defined as

$$d(g_0, g_1) = \left(\inf_{\frac{\partial}{\partial t} g_t^{-1} = -Dg_t^{-1} v_t :\, g_{t=0} = g_0, g_{t=1} = g_1} \int_{t=0}^{1} \left\| v_t(x) \right\|_L^2 dt \right)^{\frac{1}{2}} \tag{10}$$

is a metric on G. Moreover, the geodesic satisfies the Euler-Lagrange equation

$$\frac{dp_t}{dt} + (v_t \bullet \nabla) p_t + div(v_t) p_t + \sum_i \nabla((v_t)_i)(p_t)_i = 0. \tag{11}$$

Following [7], the operator L used in this paper will be in the following form

$$L = (-a\Delta + bI)^m,\ a > 0, b > 0, m \in \mathbb{N}. \tag{12}$$

3 Theory

In this paper, we try to generalize the landmark matching problem by replacing the landmarks with objects of different types. We turn to the idea of level set method for ways of representing objects. The level set method was first proposed in [10] and has been proven to be a powerful tool in front tracking as well as many other applications. For an overview on level set method, we refer the readers to [11] and the references therein.

3.1 Shape Matching

Let us start with shape matching. A shape could be represented by a level set function with the boundary of the shape being the zero level curve of the level set function (positive value inside the shape and negative outside). Throughout this paper we will use the following notation. The shapes in the template image will be denoted by the level set functions $\varphi_1, \varphi_2, \ldots, \varphi_n$ and the corresponding shapes in the study image by ψ_1, $\psi_2, \ldots, \psi_n$.

One Pair of Overlapping Shapes

In order to derive a suitable cost function that is always non-negative and only takes the value zero when the two level set functions match, we minimize the symmetric difference of the two level set functions (see [12, 13] for similar approaches).

$$D_{over}(T,S,t) = \int_\Omega H(\psi(x))[1 - H(\varphi(g_t^{-1}))]\,dx + \int_\Omega H(\varphi(g_t^{-1}))[1 - H(\psi(x))]\,dx. \tag{13}$$

The force field of this cost function is

$$f_{over}(x,t) = [1 - 2H(\psi(x))]\delta(\varphi(g_t^{-1}))\nabla\varphi(g_t^{-1}). \tag{14}$$

Here H and δ are the Heaviside and the delta function.

One Pair of Non-overlapping Shapes

The above cost function does not work for non-overlapping shapes. The reason is that by minimizing the cost function, φ will simply shrink to a point and the cost function will reach a local minimum. To overcome this, we integrate $-\psi$ in the area $\varphi>0$, $\psi<0$, and integrate $-\varphi$ in the $\varphi<0$, $\psi>0$ and we now have to initialize the level set functions to be the signed distance function to their zero level sets.

$$D_{nonoverlapping}(T,S,t) = \int_{\Omega} -\varphi(g_t^{-1})H(\psi(x))[1-H(\varphi(g_t^{-1}))]\,dx \tag{15}$$
$$+\int_{\Omega} -\psi(x)H(\varphi(g_t^{-1}))[1-H(\psi(x))]\,dx.$$

In this case the force field is given by

$$f_{nonoverlapping}(x,t) = -\{H(\psi(x))[1-H(\varphi(g_t^{-1})) \tag{16}$$
$$-\varphi(g_t^{-1})\delta(\varphi(g_t^{-1}))] + [1-H(\psi(x))]\psi(x)\delta(\varphi(g_t^{-1}))\}\nabla\varphi(g_t^{-1}).$$

3.2 Geodesic Open Curve Matching

Our task is to find a cost function between an open curve C in the template and another open curve C' in the study. We follow the idea in [14] by extending C to a closed curve (represented by the zero level set of φ_1), and we further draw a second closed curve (represented by the zero level set of φ_2), which crosses the zero level set of φ_1 only at the two end points of C. Then the open curve C is represented by

$$C = \{x \mid \varphi_1(x) = 0 \ and \ \varphi_2(x) > 0\}. \tag{17}$$

The open curve C' in the study image could also be represented by two level set functions ψ_1 and ψ_2. Let us further denote the distance functions of C and C' by $D_S(x)$ and $D_T(x)$. We then reach at the following cost function

$$D_{curve}(T,S,t) = \int_{\Omega} D_s(x)\delta(\varphi_1(g_t^{-1}))\left|\nabla\varphi_1(g_t^{-1})\right| H(\varphi_2(g_t^{-1}))dx \tag{18}$$
$$+\int_{\Omega} D_T(g_t^{-1})\delta(\psi_1(x))\left|\nabla\psi_1(x)\right| H(\psi_2(x))dx.$$

This cost function could be viewed as the sum of two line segment integrals with respect to the other curve's distance function. It can be shown that (for derivations, please refer to [15]) in this case, the force field is

$$f_{curve}(x,t) = -\frac{\delta_1}{\left|\nabla\varphi_1(g_t^{-1})\right|}\{\ H_2 < \nabla D_s, \nabla\varphi_1(g_t^{-1}) > \tag{19}$$
$$+\ D_s\delta_2 < \nabla\varphi_2(g_t^{-1}), \nabla\varphi_1(g_t^{-1}) >\}\nabla\varphi_1(g_t^{-1})$$
$$-\delta_1 D_s H_2 div\left(\frac{\nabla\varphi_1(g_t^{-1})}{\left|\nabla\varphi_1(g_t^{-1})\right|}\right)\nabla\varphi_1(g_t^{-1})$$
$$+D_s\delta_1\delta_2\left|\nabla\varphi_1(g_t^{-1})\right|\left|\nabla\varphi_2(g_t^{-1}) + \delta(\psi_1)\right|\nabla\psi_1\left| H(\psi_2)\nabla D_T(g_t^{-1}).\right.$$

Here the following short notations are used

$\delta_1 = \delta(\varphi_1(g_t^{-1}))$, $H_2 = H(\varphi_2(g_t^{-1}))$, and $\delta_2 = \delta(\varphi_2(g_t^{-1}))$.

3.3 Landmark Matching

Landmark matching problem could also be formulated in the level set framework. Let the landmark in the template be the intersection of two level set functions φ_1 and φ_2, and D_T be the distance function to the landmark. We define similarly the corresponding terms for the landmark in the study. Now consider the following cost function that recovers the sum of the distance between these two landmarks under the action of the forward and inverse mapping (see [16] for detailed discussions)

$$D_{landmark}(T,S,t) = \int_\Omega D_S \delta(\varphi_1(g_t^{-1}))\delta(\varphi_2(g_t^{-1}))\left|\nabla\varphi_1(g_t^{-1})\times\nabla\varphi_2(g_t^{-1})\right| dx + \int_\Omega D_T(g_t^{-1})\delta(\psi_1)\delta(\psi_2)\left|\nabla\psi_1\times\nabla\psi_2\right| dx \tag{20}$$

With this cost function, the force field is then

$$\begin{aligned} & f_{landmark}(x,t) = \\ & -div\left(\frac{P_{\nabla\varphi_2(g_t^{-1})}\nabla\varphi_1(g_t^{-1})}{\left|P_{\nabla\varphi_2(g_t^{-1})}\nabla\varphi_1(g_t^{-1})\right|}\left|\nabla\varphi_2(g_t^{-1})\right| D_S\right)\delta(\varphi_1(g_t^{-1}))\delta(\varphi_2(g_t^{-1}))\nabla\varphi_1(g_t^{-1}) \\ & -div\left(\frac{P_{\nabla\varphi_1(g_t^{-1})}\nabla\varphi_2(g_t^{-1})}{\left|P_{\nabla\varphi_1(g_t^{-1})}\nabla\varphi_2(g_t^{-1})\right|}\left|\nabla\varphi_1(g_t^{-1})\right| D_S\right)\delta(\varphi_1(g_t^{-1}))\delta(\varphi_2(g_t^{-1}))\nabla\varphi_2(g_t^{-1}) \\ & +\delta(\psi_1)\delta(\psi_2)\left|\nabla\psi_1\times\nabla\psi_2\right|\nabla D_T(g_t^{-1}) \end{aligned} \tag{21}$$

4 Incorporating the Level Set Method and Diffeomorphisms Generated by Infinite Dimensional Group Actions

4.1 Level Set Based Matching through Infinite Dimensional Group Actions

The level set based object matching could be integrated into the infinite dimensional group actions approach. Thus, we can examine the inexact and space time growth image matching problem as in theorem 4.1 and 4.2 of [7]. In the case of inexact matching problem, we have the following theorem

Theorem 2 (Inexact Level Set Based Object Matching)
The path that minimizes the following inexact matching problem

$$\inf_{\frac{\partial}{\partial t}g_t^{-1} = -Dg_t^{-1}v_t,\ g_{t=0}=id} \int_{t=0}^{1} \left\|v_t(x)\right\|_L^2 dt + D(T,S,t=1) \tag{22}$$

satisfies (12) and the following boundary condition at t=1

$$p(x,1)-\left[f(x,1)^T Dg_{t=1}^{-1}\right]^T=0\,. \tag{23}$$

Here $D(T,S,t)$ and f are the cost function and force field defined in section 3. For proof, simply combine the arguments in [7] and [15].

4.2 The Modified Beg's Algorithm

In this section, we will describe how to solve the inexact matching problem in 4.1 (the space-time growth problem could be handled similarly). Our algorithm is a modified version of the algorithm proposed by Faisal Beg (see [7]) for solving inexact image matching via variations with respect to the velocity field. The advantage of this approach is that we obtain the forward and inverse transformation and thus it allows us to do distance function re-initialization as described in [15].

(Modified Beg's) Algorithm for Inexact Level Set Based Object Matching

Initialize $v^{old}=0$, choose a small number ε, for all t in [0 1],

Step1 (fixed point iteration)

Solve g_t^{new} *and* g_t^{-1new} *by*

$$\frac{d}{dt}g_t^{new}=v_t^{old}(g_t^{new}),\ \frac{d}{dt}g_t^{-1new}(x)=-Dg_t^{-1new}(x)v_t^{old}(x).$$

Compute $g_1^{new}(g_t^{-1new}(\bullet))$.

Step2

$$v_t^{new}=v_t^{old}-\varepsilon\bullet grad,$$

$$grad=v_t^{old}-(L^{\dagger}L)^{-1}\left[\left|Dg_1^{new}(g_t^{-1new})\right|\sum_i f_i(g_1^{new}(g_t^{-1new}(x)),t=1)^T Dg_t^{-1new}\right]^T.$$

Set $v_t^{old}=v_t^{new}$. *Go back to step* 1 *until convergence.*

Here the algorithm is presented in the most general case in which multiple objects are to be matched. Thus, the force field has several components (f_i) and each component is defined as in section 3 depending on the nature of the object. Notice that in step 2 the gradient descent direction is in the sense of the operator $L^{\dagger}L$ instead of the usual L^2 and thus could be interpreted as a smoothing step.

We could reinitialize the distance function once in several iterations by reinitializing the level set function under the action of $g^{-1}_{t=1}$ to its zero level set and compose again with $g_{t=1}$. To be more precise, we need to add the following step in 4.2

$$Compute\ \varphi_i^*(x) \triangleq \varphi_i(g_{t=1}^{-1\,new}(x))$$

$$Reinitialize\ \varphi_i^*(x)\ to\ its\ zero\ level\ set$$

$$Update\ \varphi_i \leftarrow \varphi_i^*(g_{t=1}^{new}(x))$$

Fig. 1. Left panel: The two curves to be matched. Middle panel: The position of the curves after matching the end points. Right panel: The final result of matching the two curves.

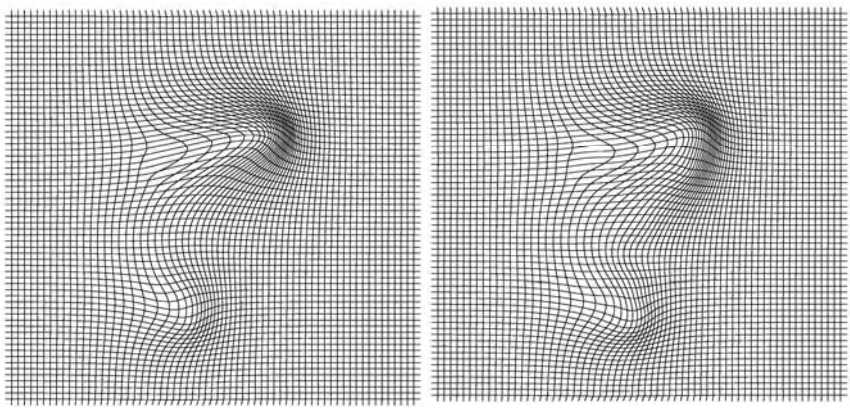


Fig. 2. Left panel: The corresponding deformation field of the middle panel in Figure 1. Right panel: The corresponding deformation field of the right panel in Figure 1.

5 Results

Open curve and landmark matching based on our method are presented (for shape matching, please refer to [15, 16]). All images are in the unit square discretized to a 64 by 64 grid, and the modified Beg's algorithm is used. The differential operator L^+L used in these two numerical examples is $50(-\Delta + id)^2$, and Fourier technique is employed to calculate the inverse of the operator in step 2 of the modified Beg's algorithm. Level set functions and distance functions in the template are re-initialized as described in 4.2 once in every100 iterations.

Figure 1 (left panel) shows the relative position of the two open curves. To test the level set based landmark matching, the proposed cost function (20) is used to match

the endpoints of the two curves and the result is shown in the middle panel of Figure 1. The corresponding deformation of the underlying grid (the geodesic path at t = 1) in this case is shown in the left panel of Figure 2. Notice that matching the endpoints alone does not ensure the matching of the whole curve. The right panel of Figure 1 shows the matching of the whole curve by minimizing cost function (18), with the corresponding deformation (the geodesic path at t = 1) shown in the right panel of Figure 2. For comparison, Figure 3 shows the geodesic path in these two examples at time $t = 0.5$. The length of the geodesic path is 8.97 in the case of endpoint matching, and 10.78 in the case of whole curve matching.

This paper was partially supported by NIH grant NS30308 and P20MH65166.

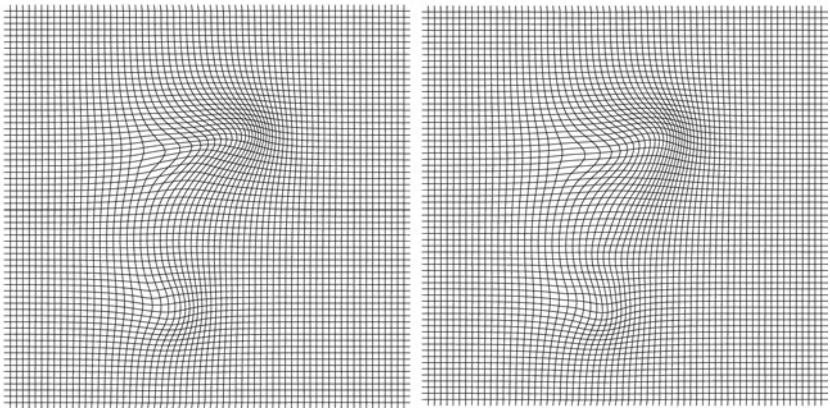


Fig. 3. Left panel: The corresponding geodesic path at $t = 0.5$ of the middle panel in Figure 1. Right panel: The corresponding geodesic path at $t = 0.5$ of the right panel in Figure 1.

References

1. U. Grenander and M. I. Miller, "Computational anatomy: An emerging discipline," *Quarterly of Applied Mathematics*, vol. 56, pp. 617-694, 1998.
2. P. Thompson and A. W. Toga, "A framework for computational anatomy," *Computing and Visualization in Science*, vol. 5, pp. 13-34, 2002.
3. R. Bajcsy and S. Kovacic, "Multiresolution Elastic Matching," *Computer Vision Graphics and Image Processing*, vol. 46, pp. 1-21, 1989.
4. G. E. Christensen, R. D. Rabbitt, and M. I. Miller, "Deformable templates using large deformation kinematics," *IEEE Transactions on Image Processing*, vol. 5, pp. 1435-47, 1996.
5. G. E. Christensen, R. D. Rabbitt, and M. I. Miller, "3D brain mapping using a deformable neuroanatomy," 1994.
6. G. E. Christensen, S. C. Joshi, and M. I. Miller, "Volumetric transformation of brain anatomy," *IEEE Transactions on Medical Imaging*, vol. 16, pp. 864-77, 1997.
7. M. I. Miller, A. Trouve, and L. Younes, "On the metrics and Euler-Lagrange equations of computational anatomy," *Annual Review of Biomedical Engineering*, vol. 4, pp. 375-405, 2002.

8. P. Dupuis, U. Grenander, and M. I. Miller, "Variational problems on flows of diffeomorphisms for image matching," *Quarterly of Applied Mathematics*, vol. 56, pp. 587-600, 1998.
9. D. Mumford, "Pattern Theory: the Mathematics of Perception," in *ICM 2002*, vol. 3, 2002.
10. S. Osher and J. A. Sethian, "Fronts Propagating with Curvature-Dependent Speed - Algorithms Based on Hamilton-Jacobi Formulations," *Journal of Computational Physics*, vol. 79, pp. 12-49, 1988.
11. S. Osher and R. P. Fedkiw, "Level set methods: An overview and some recent results," *Journal of Computational Physics*, vol. 169, pp. 463-502, 2001.
12. R. T. Whitaker, "A level-set approach to image blending," *IEEE Transactions on Image Processing*, vol. 9, pp. 1849-61, 2000.
13. N. Paragios, M. Rousson, and V. Ramesh, "Matching Distance Functions: A Shape-to-Area Variational Approach for Global-to-Local Registration," *ECCV*, 2002.
14. P. Smereka, "Spiral crystal growth," *Physica D*, vol. 138, pp. 282-301, 2000.
15. W. H. Liao, A. Jhuu, M. Bergsneider, L. A. Vese, S. C. Huang, and S. Osher, "From Landmark Matching to Shape and Open Curve Matching: A Level Set Approach," in *UCLA CAM report*, 2002.
16. W. H. Liao, "Mathematical techniques in object matching and computational anatomy: a new framework based on the level set method," in *Biomathematics*. Los Angeles: UCLA, 2003.

Constructing Data-Driven Optimal Representations for Iterative Pairwise Non-rigid Registration

Stephen Marsland and Carole J. Twining

Imaging Science and Biomedical Engineering
University of Manchester
Manchester M13 9PT, U.K.
{stephen.marsland,carole.twining}@man.ac.uk

Abstract. Non-rigid registration of a pair of images depends on the generation of a dense deformation field across one of the images. Such deformation fields can be represented by the deformation of a set of knotpoints, interpolated to produce the continuous deformation field. This paper addresses the question of how best to choose the knotpoints of such a representation based on all of the available image information. These knotpoints are not landmarks, they can be positioned anywhere in the images, and do not necessarily correspond to any image feature. We use an iterative, data-driven algorithm for the selection of knotpoints, and a novel spline that interpolates smoothly between knotpoints. The algorithm produces a low-dimensional representation of the deformation field that can be successively refined in a multi-resolution manner. We demonstrate the properties of the algorithm on sets of 2D images and discuss the extension of the algorithm to 3D data.

1 Introduction

Non-rigid registration algorithms deform a dense set of pixels or voxels in an image so that structures in one image can be more easily matched with those in another, reference image. Such algorithms are particularly useful for medical images, since they enable tasks as diverse as the correction for patient motion (e.g., [15]), the integration of multi-modal (e.g., MR, PET and CT) images for either a single patient or multiple patients [17], and automated segmentation via registration onto an anatomical atlas [6].

The registration of a pair of images is described in terms of a deformation field, which encodes information about the variability of structures in the images. Typically, non-rigid registration algorithms represent the deformation field using splines [11,14,15], variational methods [4], or physical models such as elastic matching [6] or fluid flow [3]. Whatever the method of representing the deformation field, the algorithms all attempt to optimise some similarity measure that evaluates the degree to which two images match by varying the model parameters that control the deformation field. Common similarity measures include voxel intensity measures such as sum-squared difference and statistical measures

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 50–60, 2003.

such as Mutual Information [9,17] and the correlation ratio [12]. The benefits of the last two are that they enable the registration of images obtained using different imaging modalities.

Of particular relevance to the work described in this paper are the Free-Form Deformations (FFD) used by Rueckert et al. [15,16,7] and Rohlfing and Maurer [13], which are based on B splines, and methods based on thin-plate splines, such as Rohr et al. [14] and Meyer et al. [11].

Free-form deformations are based on smoothing (as opposed to interpolating) splines, whose control points lie on a regular grid. Implementations can involve a sequence of global coarse to fine grids [15], refinement over a regular region [13], or simulating a non-uniform control point distribution using data-driven methods [16]. However, in all these cases, control points are still constrained to the nodes of some regular lattice, which means that useful registration points that lie between nodes may be missed. B splines are local functions, and ensuring local consistency at the edges of the refined section of the grid, or between grids at different resolutions, is non-trivial and requires a computationally expensive algorithm such as the Oslo algorithm [5].

Methods based on thin-plate splines are inspired by work on morphometrics, see for example [2]. In [11] a thin-plate spline is used to interpolate between the knotpoints, which are placed on both images by hand and then optimised to improve the mapping. Rohr [14] adapts the method in two ways, first by amending the thin-plate spline to be a smoothing rather than interpolating spline so that the curve does not need to go exactly through the knotpoints, and second by producing a 'semi-automatic' method of selecting landmarks (a similar 'semi-automatic' method using B splines is described in [7]), both methods rely on user-selected landmark positions. We contend that the thin-plate spline is not ideal for the task of warp interpolation. The spline is only asymptotically flat, and positions that are between knotpoints may be forced to move significantly further than the knotpoints. This is demonstrated in Figs. 1 and 2 in section 2.

This paper introduces a novel method of pairwise non-rigid registration that focusses particularly on a data-driven method of selecting knotpoints for both the initialisation of the algorithm and to iteratively increase the resolution of the registration. Selecting knotpoints based on the information in the images means that the dimensionality of the representation is kept low, as knotpoints are only placed where they are needed – in particular, features such as the background, which do not change, will tend not to get knots and are therefore not actively warped. The clamped-plate spline (CPS), introduced in [10] and described in more detail in section 2, is used to interpolate the warp between knotpoints. The CPS deals efficiently with additional knotpoints added during subsequent iterations of the algorithm and has strong boundary conditions. The interpolation between knotpoints is smooth, and decreases with distance from the knotpoints (see Fig. 2).

We consider a variety of methods of choosing new knotpoints; those based on information in only the reference image (e.g., image features such as edges or corners) and those that use information from both images, in particular, discrepancies in the objective function. We compare the results of using these

methods with the naïve approach of putting a lattice of knotpoints across the image, both in terms of computational efficiency and in terms of the quality of the registration produced, using images of axial and coronal slices of the human head. In this paper the example images are all two-dimensional, since it is much simpler and faster to run the experiments on 2D images. Furthermore, it is easier to investigate the basic properties and the results of using the algorithm in 2D. Extending the method to 3D is relatively simple, it involves no new conceptual steps, and the particular spline basis used is readily extensible to 3D – this is given in the paper.

2 The Interpolating Clamped-Plate Spline

We consider a vector-valued spline function $\boldsymbol{f}(\boldsymbol{x}),\ \boldsymbol{x} \in \mathbb{R}^n$ that interpolates between data values at a set of knotpoints $\{\boldsymbol{x}_i : i = 1 \text{ to } N\}$, where $\boldsymbol{f}(\boldsymbol{x}_i) = \boldsymbol{f}_i$. We restrict ourselves to the class of interpolating splines, which can be expressed as the minimiser of a functional Lagrangian of the form:

$$E[\boldsymbol{f}] = \int_{\mathbb{R}^n} d\boldsymbol{x} \, \|L\boldsymbol{f}(\boldsymbol{x})\|^2 + \sum_{i=1}^{N} \lambda_i \left(\boldsymbol{f}(\boldsymbol{x}_i) - \boldsymbol{f}_i\right), \tag{1}$$

where L is some scalar differential operator. The first term in the Lagrangian is the smoothing term; the second term with the Lagrange multipliers $\{\lambda_i\}$ ensures that the spline defined by the functional minimisation exactly fits the data at the knotpoints. The general solution of this is of the form:

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{g}(\boldsymbol{x}) + \sum_{i=1}^{N} \boldsymbol{\alpha}_i G(\boldsymbol{x}, \boldsymbol{x}_i), \tag{2}$$

where:

$$\text{the affine function } g \text{ is a solution of: } L\boldsymbol{g}(\boldsymbol{x}) = 0, \tag{3}$$

$$\text{the Green's function } G \text{ is a solution of: } \left(L^\dagger L\right) G(\boldsymbol{x}, \boldsymbol{y}) \propto \delta(\boldsymbol{x} - \boldsymbol{y}), \tag{4}$$

and $L^\dagger$ is the Lagrange dual of L. The choice of operator L and boundary conditions then define a particular spline basis – for example, if L is the biharmonic differential operator $L = -\triangle = -\nabla^2$ then the first term of equation (1) is the approximate Willmore energy (i.e., the thinplate bending energy) of the displacement fields. In this case the Green's function:

$$G_{TPS}(\boldsymbol{x}, \boldsymbol{y}) = -|\boldsymbol{x} - \boldsymbol{y}|^2 \log |\boldsymbol{x} - \boldsymbol{y}|^2 \tag{5}$$

produces the thin-plate spline interpolant. An alternative Green's function is the polyharmonic clamped-plate [1], which satisfies both Dirichlet and von Neumann boundary conditions, so that the function is identically zero on and outside the unit ball in n dimensions. The clamped-plate Green's function is defined for all

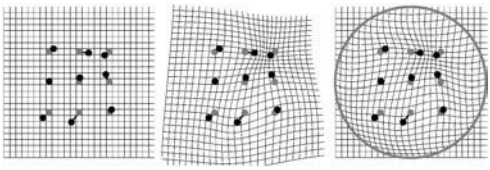


Fig. 1. *Left:* Grid with initial (grey) and final (black) knotpoint positions. *Middle:* TPS interpolated warp. *Right:* CPS interpolated warp, showing the bounding unit circle.

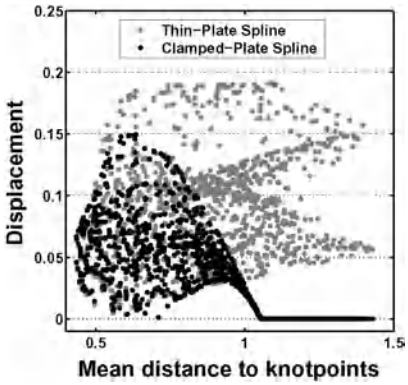


Fig. 2. Displacement (TPS grey, CPS black) as a function of the mean distance to knotpoints for the warp shown in Fig. 1.

operators $L^\dagger L = (-\nabla^2)^m$ and dimensionalities n, but in particular we give the biharmonic operator in 2 dimensions (G_2^2) and the triharmonic operator in 3 dimensions (G_3^3); for further details see [18]:

$$\left.\begin{aligned} G_2^2(\boldsymbol{x},\boldsymbol{y}) &= \|\boldsymbol{x}-\boldsymbol{y}\|^2\left(\tfrac{1}{2}(A^2-1)-\log A\right) \\ G_3^3(\boldsymbol{x},\boldsymbol{y}) &= \|\boldsymbol{x}-\boldsymbol{y}\|\left(A+\tfrac{1}{A}-2\right) \end{aligned}\right\} A(\boldsymbol{x},\boldsymbol{y}) = \frac{\sqrt{|\boldsymbol{x}|^2|\boldsymbol{y}|^2-2\boldsymbol{x}\cdot\boldsymbol{y}+1}}{\|\boldsymbol{x}-\boldsymbol{y}\|}. \tag{6}$$

The differences between the thin-plate (TPS) and clamped-plate (CPS) splines for interpolation of deformations can be clearly seen in Figs. 1 and 2. The warp is interpolated from the movement of a small number of knotpoints, where in this example the displacements are chosen so that the affine part of the TPS warp is identically zero. The TPS warps the entire grid because it has weak boundary conditions – the Green's function (equation 5) decays logarithmically, and hence the warp is only *asymptotically* flat – by contrast the CPS only warps that part of the grid inside the unit circle. Fig. 2 shows the displacement of sample points against their mean distance from the set of knotpoints, for a set of random knotpoint displacements. It can be seen that the CPS interpolant is more localised, with the interpolated displacements of sample points decaying as the distance from the knotpoints increases.

2.1 Using the CPS for Image Registration

Given a set of knotpoints defined on an image, and their displacements, the CPS interpolates this warp in a smooth and localised fashion across the whole image. The local property of the CPS means that adjusting the displacement of a given knotpoint tends to effect only regions in the vicinity of that knotpoint, leaving unaffected possibly good registrations of other areas of the image. This

property will be seen to be important when we consider the optimisation of knotpoint displacements. The original knotpoint positions are defined on the reference image of a pair, and the final positions on the second (free) image. These knotpoints, and the warp that they generate, then define a dense correspondence between the pair of images.

Notation. We consider an image I to be a scalar-valued function defined on some dense point set X_0. Initially, these points will be a regular grid (the set of pixel or voxel centres). By interpolation we can extend such an image to be continuously defined within the hull of X_0; we can also extrapolate, defining values of I outside this hull as either the background value or the median value of $I(X_0)$, as appropriate.

After applying a warping function $h : X_0 \to X_h = h(X_0)$ to the grid, where the warp could be a general affine warp g, a CPS warp f, or some composition of the two, $h = f \circ g$, the warped grid is then used to generate the free image $I^{(h)}$ using the pullback map h^*:

$$h^* : I \to I^{(h)}, \text{ where } I^{(h)}(X_0) \equiv I(h(X_0)), \Rightarrow I^{(h)}(h^{-1}(X_0)) \equiv I(X_0), \quad (7)$$

which requires interpolation of the function I from the points X_0 to $h(X_0)$. The CPS warps are defined by the initial and final positions $Q_0 = \{\boldsymbol{q}_{0\beta} \in \mathbb{R}^n, \beta = 1, \ldots, n_k\}, Q_1 = \{\boldsymbol{q}_{1\beta}\}$ of the set of knotpoints, which are first defined on the reference image I_r as Q_{orig} and then affinely warped, $Q_0 = g(Q_{\text{orig}})$. The non-rigid CPS warp generated by these set of knots is denoted by $f(\cdot) = \omega(Q_0, Q_1)$, with the action on points $\boldsymbol{x} \in \mathbb{R}^n$ being:

$$f(\boldsymbol{x}) = \boldsymbol{x} + \sum_{\beta=1}^{n_k} \boldsymbol{\alpha}_\beta G_n^m(\boldsymbol{x}, \boldsymbol{q}_{0\beta}), \quad (8)$$

where the vector-valued coefficients $\{\boldsymbol{\alpha}_\beta\}$ are found by solving the exact matching conditions for the knotpoints $f(Q_0) = Q$ and G_n^m is the m^{th} order polyharmonic clamped-plate Green's functions defined in equation (6).

The similarity between two images I and I' is quantified by the evaluation of some objective function $\boldsymbol{\Phi}\left(I(X_0), I'(X_0)\right)$. In the implementation presented in this paper we use Mutual Information (MI), so that the similarity between two images is [9]:

$$\boldsymbol{\Phi}_{H_S}(I, I') = H_S(I) + H_S(I') - H_S(I, I'), \quad (9)$$

where $H_S(I) = -\sum_i p_i \log p_i$ is the Shannon entropy. Warps are selected to minimise the difference between the free image I and the reference image I_r:

$$g' = \arg\min_g \boldsymbol{\Phi}\left(I_r, I^{(g)}\right), \; f = \arg\min_f \left[\boldsymbol{\Phi}\left(I_r, I^{(f\circ g)}\right) + \lambda E[f]\right], \quad (10)$$

where $\lambda E[f]$ is a regularisation term proportional to the energy of the warp. The optimised final knotpoint positions on the free image I are then given by:

$$Q_1 = \arg\min_Q \left[\boldsymbol{\Phi}\left(I_r(X_0), I^{(\omega(Q_0,Q)\circ g)}(X_0)\right) + \lambda E[\omega(Q_0, Q)]\right]. \quad (11)$$

3 Data-Driven Knotpoint Selection

In our algorithm, knotpoints for the spline are chosen based on both the images. Knotpoints are placed on the reference image and their positions on the free image drive the optimisation. The algorithm is iterative, with knotpoints being added to the spline during each iteration of the algorithm based on a comparison of the reference image, which does not change, and the current appearance of the free image. An initial set of knotpoints are chosen based on the affinely aligned reference and free images, and an initial non-rigid registration is computed by optimising over their positions on the free image. The reference image (which remains unchanged) and the newly-warped free image are then used to choose where the next set of knotpoints should be placed.

In this section we describe the methods of selecting knotpoints that we test in this paper. We particularly focus on using the discrepancies in the objective function to select the new knotpoint positions, since it is optimising the objective function that drives the registration. The form of this discrepancy function depends on the objective function, and for particular choices of the objective function, will correspond with some of the 'joint image pair measures' mentioned in [16]. A specific instance of our method, based on Mutual Information, is described in section 3.1. Although some of the knotpoint positioning strategies may be similar to those used to select landmarks, it does not mean that this method of pairwise non-rigid registration is related to landmark-based methods of rigid or non-rigid registration (e.g., [14,2]). Such landmark-based methods rely on finding significant *corresponding* landmarks on all images, and the final registration is only as good as this initial landmark correspondence.

Our method relies on selecting knotpoints on the reference image using information from both images. Knotpoints are *not* landmarks – they can be positioned anywhere in the image, and need not correspond with any structure. In our method a poor initial choice of knotpoint is not a major problem – although such an occurrence will delay convergence, it will be iteratively corrected.

3.1 Generating the Discrepancy Function from the Objective Function

This section describes a discrepancy function for the case where the objective function is Mutual Information; a different choice of objective function (e.g., normalised Mutual Information) would generate a different discrepancy function. The computation of the Mutual Information requires the computation of the joint histogram of pixel intensities in the two images [9]. We use this computation to estimate the distribution of pixel intensities in the reference image that correspond to pixels of a particular intensity in the free image, and model the distribution as a Gaussian. We can then identify those pixels that are outliers to this distribution, that is, the positions on the reference image where the probability of getting that particular pixel intensity are low, given the free image. This identifies pixels on the reference image that are currently mis-registered.

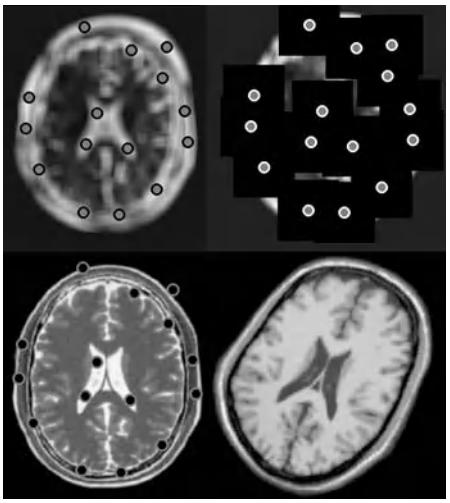


Fig. 3. *Bottom:* The reference and free images. *Top:* The MI-based discrepancy between the reference and affinely-aligned free images, showing knotpoints and exclusion zones.

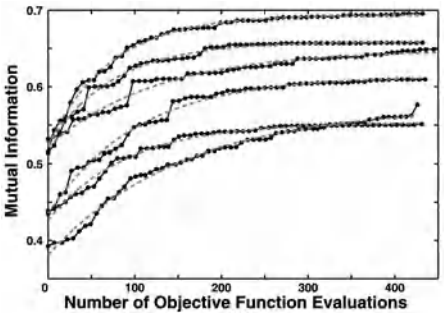


Fig. 4. Exponential fits (dashed lines) to the MI curves for 6 pairs of registrations.

One example of such a multi-modality discrepancy image is shown in Fig. 3 [1]. Once a knotpoint has been placed, we then define a square *exclusion zone* around this point, to prevent each area of misalignment being too densely covered with knotpoints. The size of this exclusion zone hence defines the scale of this particular step of the registration. As will be shown later, the edges of structures larger than this scale (i.e., the skull) will be precisely aligned (provided enough knotpoints are defined), whereas the alignment of edges of smaller structures (e.g., the ventricles) will be only approximate until a finer scale of the registration is reached.

3.2 Comparing Methods of Knotpoint Selection

We compare several methods of choosing knotpoints:

Points on a Grid. The FFDs described in section 1 require that the knotpoints are defined on a grid. We compare this method for choosing the initial set of knotpoints by laying a regular grid of knotpoints across the image.

Edges. The Canny edge detector was used to highlight strong edges in the image. A set of rays spaced at equal angles from the centre of the image was then constructed. For each ray the strong edge that is furthest from the centre was selected as a knotpoint. This tends to place evenly-spaced knotpoints around

[1] Simulated T1 and T2 slices from BrainWeb: `http://www.bic.mni.mcgill.ca/brainweb/`. The free image has been rotated and artificially warped.

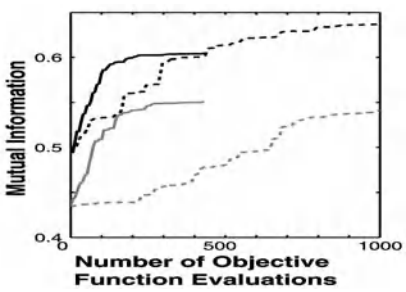


Fig. 5. Convergence of MI for two pairs of axial slices, using 16 knotpoints based on MI discrepancy (solid lines) versus a regular 10×10 grid of knotpoints (dashed lines).

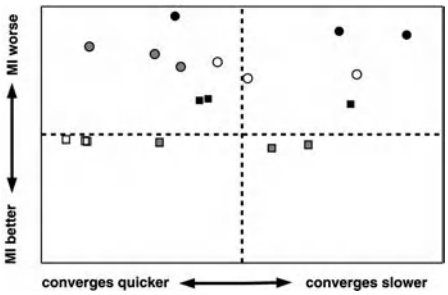


Fig. 6. Final MI against speed of convergence for 6 registrations (6 shapes and colours of plotted points). The dotted lines show the MI discrepancy method of selecting knotpoints, against which the other methods (16 skull points, 10 skull points and a grid of 16 points) were standardised.

the outer edge of an isolated object in the image, for example, around the skull in images of a brain.

4 Experiments

Our main data set consists of 12 pairs of 2D T1-weighted MR images of axial brain slices. The 227×227 pixel images were from 24 different normal individuals, and the slices were chosen to show the anterior and posterior parts of the lateral ventricles. We also used images of coronal brain slices from normal individuals. The first task is to show that our algorithm does indeed converge in a reasonable time. To evaluate this, we choose 16 initial knotpoints according to the MI-discrepancy for each of the pair of axial slices. To optimise the objective function, we adjust the final position of a single knotpoint at a time. We loop over the set of knotpoints in a random order, and repeat several times. An example initial registration is shown in Fig. 7. In Fig. 4 we show the variation of the Mutual Information as a function of the number of objective function evaluations. For the sake of clarity, we show only 6 of the 12 plots, chosen so that they show the full span of the set. As can be seen from the figure, the optimisation converges within a few hundred function evaluations for all pairs of the set, and the convergence of the Mutual Information is fitted extremely well by exponential decay curves. We also compared these MI-based methods of knotpoint selection against regular grids of knotpoints. In Fig. 5 we show the convergence of the Mutual Information for two pairs of axial images, using 16 knotpoints selected according to the MI discrepancy, as well as 100 knotpoints on a regular grid. It can be seen that the grid of knotpoints converges much more slowly, with a relatively small gain in the final value of the Mutual Information. In Fig. 6 we show a quantitative comparison of discrepancy-based selection methods versus

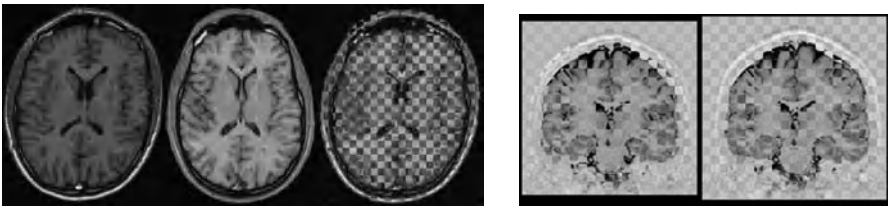


Fig. 7. Example initial registration using 16 knotpoints selected using MI discrepancy of a pair of axial slices and a pair of coronal slices. From the left, the reference, unwarped free, and chequerboard registered images for the axial images, and the chequerboard affinely registered and non-rigidly registered images for a pair of coronal images. Note the extremely good registration of the skull and brain edges in both cases.

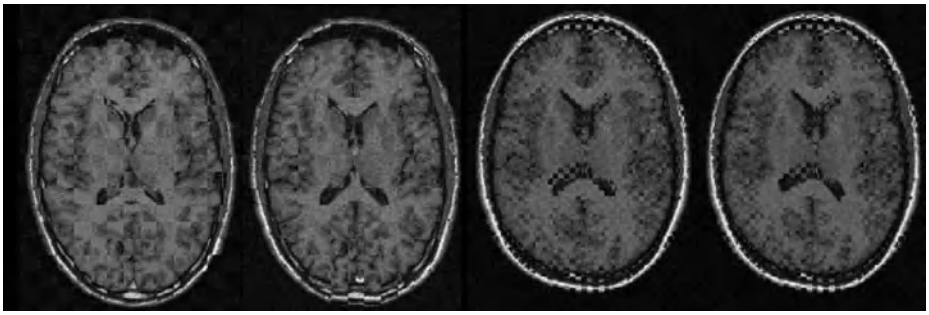


Fig. 8. Chequerboard images of registered pairs showing refinement of the registration. *Left* 10 and (*right*) 16 knotpoints were used for the initial registrations, and 10 knotpoints added for the next step. Note the improved alignment of parts of the ventricular structure.

edge-based skull points and small grids, for 6 of the 12 pairs of axial images. It can be seen that the discrepancy-based selection method tends to give better final values of the Mutual Information, whilst no other method gives faster convergence over all the examples.

4.1 Adding Additional Knotpoints

Once a set of initial knotpoints has been found and the final positions optimised, the next step is to increase the resolution of the deformation field. This is done by decreasing the size of the exclusion zones around the current knotpoints. The discrepancy between the reference image and the current free image is then used to select new, additional knotpoints, as in section 3. As regards optimisation, it was found that a full optimisation was only required over final positions of the new knotpoints; the final positions of the previous set of knotpoints required only minimal re-optimisation. An example of the refinement of the registration is shown in Fig. 8; note that the registration of the positions of the ventricles has improved, without spoiling the previous good registration of the skulls.

5 Discussion and Conclusions

This paper has introduced a method of non-rigid registration that generates a low-dimensional, data-driven representation of the deformation field using a small number of knotpoints selected in an iterative data-driven way, and a novel spline. We have demonstrated this iterative algorithm on pairwise registration of 2D T1-weighted MR images of axial and coronal slices of the human head, with all 12 pairs of brains being registered successfully. The paper has shown that selecting knotpoints using the discrepancy in the Mutual Information objective function generates low-dimensional representations of the deformation fields that converge with comparable speed to methods based on either a regular grid of knotpoints or knotpoints placed on strong edges in the image.

The algorithm that we have described is an iterative one, with additional knotpoints being added at each iteration in order to refine the registration. We have shown two iterations of the algorithm, with 16 knotpoints in the first iteration and 10 more added in the second iteration, producing reasonable registrations; more accurate registrations would be produced by further iterations. These additional knotpoints are added in those places where the MI objective function still shows discrepancies between the reference and free images. The registrations that we have shown in this paper have been 2D, as we have focussed on demonstrating the effects of the new method. The spline basis is also given for 3D in the paper. All the results shown in this paper were produced using a MATLAB implementation of the algorithm, and one iteration of the algorithm on a pair of images took approximately an hour on a 1.8GHz PC.

One of the motivations of our algorithm is groupwise registration, where a set of images are aligned so that they have a common frame of reference. This facilitates the modelling of variations in structures between and within disease groups and normals. However, useful analysis requires a common low-dimensional representation of the warps. Work to date has either used the densely-sampled deformation vectors directly [8], or employed a smooth, continuous representation of them [15]. Our data-driven algorithm can be extended to generate a common low-dimensional representation by choosing new knotpoints based on the errors in the objective function taken over the entire group of images. This extension to groupwise registration, and the analysis of the deformation fields produced, is currently being investigated.

Acknowledgements

This research was supported by the MIAS IRC project, EPSRC grant number GR/N14248/01.

References

1. T. Boggio. Sulle funzioni di green d'ordine m. *Rendiconti - Circolo Matematico di Palermo*, 20:97–135, 1905.
2. F. L. Bookstein. *Morphometric tools for landmark data.* CUP, 1991.

3. M. Bro-Nielsen and C. Gramkow. Fast fluid registration of medical images. In *Proceedings of Visualization in Biomedical Computing (VBC)*, pages 267–276, 1996.
4. C. Chefd'Hotel, G. Hermosillo, and O. Faugeras. A variational approach to multimodal image matching. In *Proceedings of IEEE Workshop on Variational and Level Set Methods (VLSM'01)*, pages 21 – 28, 2001.
5. D. R. Forsey and R. H. Bartels. Hierarchical B-spline refinement. *ACM Transactions in Computer Graphics*, 22(4):205–212, 1988.
6. J. Gee, M. Reivich, and R. Bajcsy. Elastically deforming 3D atlas to match anatomical brain images. *Journal of Computer Assisted Tomography*, 17(2):225–236, 1993.
7. T. Hartkens, D.L.G. Hill, A.D. Castellano-Smith, D.J. Hawkes, C.R. Maurer Jr., A.J. Martin, W.A. Hall, H. Liu, and C.L. Truwit. Using points and surfaces to improve voxel-based non-rigid registration. In *Proceedings of MICCAI 2002*, number 2489 in Lecture Notes in Computer Science, pages 565 – 572, 2002.
8. L. LeBriquer and J. Gee. Design of a statistical model of brain shape. In *Proceedings of IPMI'97*, Lecture Notes in Computer Science 1230, pages 477–482, 1997.
9. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187 – 198, 1997.
10. S. Marsland and C. J. Twining. Clamped-plate splines and the optimal flow of bounded diffeomorphisms. In *Statistics of Large Datasets, Proceedings of Leeds Annual Statistical Research Workshop*, pages 91–95, 2002.
11. C. R. Meyer, J. L. Boes, B. Kim, P. H. Bland, K. R. Zasadny, P. V. Kison, K. Koral, K. A. Frey, and R. L. Wahl. Demonstrations of accuracy and clinicial versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations. *Medical Image Analysis*, 1(3):195 – 206, 1997.
12. A. Roche, G. Malandain, X. Pennec, and N. Ayache. The correlation ratio as a new similarity measure for multimodal image registration. In *Proceedings of MICCAI'98*, pages 1115 – 1124, 1998.
13. T. Rohlfing and C.R. Maurer, Jr. Intensity-based non-rigid registration using adaptive multilevel free-form deformation with an incompressibility constraint. In *Proceedings of MICCAI 2001*, Lecture Notes in Computer Science 2208, pages 111 – 119, 2001.
14. K. Rohr, H. S. Stiehl, R. Sprengel, T. M. Buzug, J. Weese, and M. H. Kuhn. Landmark-based elastic registration using approximating thin-plate splines. *IEEE Transactions on medical imaging*, 20(6):526–534, 2001.
15. D. Rueckert, A. F. Frangi, and J. A. Schnabel. Automatic construction of 3D statistical deformation models using non-rigid registration. In *Proceedings of MICCAI'01*, volume 2208 of *Lecture notes in Computer Science*, pages 77–84, 2001.
16. J. A. Schnabel, D. Rueckert, M. Quist, J. M. Blackall, A. D. Castellano-Smith, T. Hartkens, G. P. Penney, W. A. Hall, H. Liu, C. L. Truwit, F. A. Gerritsen, D.L.G. Hill, and D. J. Hawkes. A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations. In *Proceedings of MICCAI 2001*, number 2208 in Lecture Notes in Computer Science, pages 573 – 581, 2001.
17. C. Studholme, D. Hill, and D. Hawkes. Automated three-dimenensional registration of magnetic resonance and positron emission tomography by multiresolution optimisation of voxel similarity measures. *Medical Physics*, 24(1):25 – 35, 1997.
18. C. J. Twining and S. Marsland. Constructing diffeomorphic representations of non-rigid registrations of medical images. In *Proceedings of IPMI, to appear*, 2003.

Non-rigid Registration of Serial Intra-operative Images for Automatic Brain Shift Estimation

Valerie Duay, Tuhin K. Sinha, Pierre-François D'Haese, Michael I. Miga, and Benoit M. Dawant

Vanderbilt University, Nashville, TN, 37235, USA
http://www.vuse.vanderbilt.edu/~mip-web

Abstract. Measurement of intra-operative brain motion is important to provide boundary conditions to physics-based deformation models that can be used to register pre- and intra-operative information. In this paper we present and test a technique that can be used to measure brain surface motion automatically. This method relies on a tracked laser range scanner (LRS) that can acquire simultaneously a picture and the 3D physical coordinates of objects within its field of view. This reduces the 3D tracking problem to a 2D non-rigid registration problem which we solve with a Mutual Information-based algorithm. Results obtained on images of a phantom and on images acquired intra-operatively that demonstrate the feasibility of the method are presented.

1 Introduction

Image-guided surgery aims at bringing pre-operative information to the surgeon during the procedure. Most often, this involves registering pre-operative images with the patient in the OR. A number of methods have been developed for this purpose but until recently these have involved rigid body registration techniques. Although rigid body techniques have proven clinically useful, there is a body of literature that shows that brain deforms during the procedure [1,2,3]. When this is the case, rigid body transformations are not sufficient to register accurately pre- and intra-operative information. This has lead several research groups to develop methods and techniques that can compensate for intra-operative brain shift. These methods fall into two broad categories. The first involves intra-operative imaging (for instance interventional MR [4,5], ultrasound [6,7], or CT [8]). When available, intra-operative images can be registered to the pre-operative images using a number of non-rigid intra- or inter-modal registration methods developed over the years (see for instance Meyer et al. [9], Rueckert et al. [10]). Although attractive, this solution is only possible at a few sites that have the required imaging equipment. As an alternative, others have proposed to use physical models [11,12]. Displacements measured at the surface of the brain can then be propagated through the entire volume based on these models. Surface displacements can be measured with a tracked probe [13]

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 61–70, 2003.

or, as is the case at our institution, a tracked laser range scanner (LRS) [14]. When a tracked probe is used, fiducial points such as the intersection between sulci or vessels' branching points need to be identified at the brain surface and their position tracked over time intra-operatively. This is not only tedious but also error prone. Laser range scanners on the other hand offer the possibility to track points on the brain surface automatically and thus estimate their 3D displacement automatically as well. A laser range scanner as the one we use permits the simultaneous acquisition of the x, y, and z physical coordinates of the objects within its field of view and of a 2D RGB picture of these objects. The scanner also provides a mapping between each point whose physical coordinates are acquired and its position in the RGB image. If a series of images with their associated 3D physical coordinates can be acquired and if a correspondence between the pixels in these images can be established, the temporal displacement of any point within the field of view of the LRS can be computed. Thus, tracking brain surface deformation only requires establishing a correspondence between points in a series of 2D RGB images which can be cast as a registration problem. Clearly, as the surgical procedure evolves, the surface of the brain changes. At times these changes may be drastic, e.g., when a resection is made to remove a tumor close to the brain surface or when an incision is made to access a deeper one. To accommodate these changes non-rigid registration methods are called for. One possibility would be to identify homologous points in the images to be registered and compute a transformation between these images based on these points. This approach is not automatic and thus undesirable. In this paper, we propose and evaluate a method that permits the automatic registration of these images. In the remainder of the paper we present the method we have used as well as results we have obtained both with phantom and real images acquired intra-operatively.

2 Methods

2.1 Data Acquisition

The data used in this study has been acquired with a laser-range scanning device (RealScan3D USB, 3D Digital Corp, Bethel, CT, USA). This scanner is mounted on a vibration-damped monopod that is brought into and out of the surgical field-of-view (SFOV) manually. A thorough discussion of the range scanner can be found in [14]. Intra-operative images are acquired as follows. After dural opening, the monopod and scanner are brought into the SFOV and the laser scanning extents (left and right margins) are calibrated to cover the width of the craniotomy. A laser stripe is then passed over the brain's surface and range data is collected using the principle of optical triangulation. After acquisition, the scanner and monopod are moved out of the SFOV. The entire data acquisition process adds approximately 1.5 minutes per scan to the operating time and has been approved for clinical use by the Vanderbilt University IRB (VUIRB). A 480x640 pixels RGB bitmap image registered to the range data is acquired at the time of scanning.

2.2 Serial Image Registration

Prior to registration, the RGB images are transformed into gray level images. These images are then first registered using a rigid body transformation (three degrees of freedom: rotation and translations in the horizontal and vertical directions). To do so we use a Mutual Information-based method as proposed by Maes et al. [15]. Our implementation of this algorithm permits registration at multiple resolutions. Here we use two levels (240x320 and 480x640 pixels). We estimate the probability density functions required for the computation of the Mutual Information from the joint histogram of the images. We use 32 bins to build these histograms. In a second step, we refine the results obtained after rigid body registration with a non-rigid registration algorithm we have recently proposed (Rohde et al. [16]). This method has been inspired by the work of Meyer et al. [9] and Rueckert et al. [10]. In this approach, the deformation field that registers one image to the other is modeled with a linear combination of radial basis functions with finite support. The similarity measure used to drive the registration process is the Mutual Information between the images. Our method differs from others in several ways. In our approach we can work on an irregular gird, we adapt the compliance of the transformation locally, we optimize our transformation sequentially on separate regions to speed up the process, and we have derived constraint schemes on the transformation coefficients to enforce the topological correctness of the transformation. Space precludes giving detailed information on this algorithm but these details can be found in [16]. Our algorithm computes the final deformation field iteratively across scales and resolutions (in this context, resolution means the spatial resolution of the image while the scale is related to the transformation itself). A standard image pyramid is created to apply the algorithm at different resolutions. At each resolution, the scale of the transformation is adapted by modifying the region of support and the number of basis functions. The final deformation field is computed as the sum of deformation fields computed at a series of levels, with one level referring to a particular combination of scale and resolution. For the images presented herein, we have used two resolutions (120x160 and 240x320 pixels). At the lowest resolution, we used 3 transformation scales (basis functions with region of support ranging from 40 pixels to 24 pixels). At the higher resolution, we use 5 transformation scales (basis functions with region of support ranging from 30 pixels to 8 pixels). As was the case for the rigid body registration algorithm, the probability density functions required for the computation of the Mutual Information are estimated from the joint histograms built with 32 bins.

2.3 Data Sets Used in This Study

We have tested our approach on one phantom and three *in vivo* cases. A silicon impression of a cortical surface mold was used as the scanning phantom. This phantom was placed in a clamp and scanned three times. The first time the phantom was not compressed, the second it was compressed from the top, and the third it was compressed from the top and the bottom. The three *in vivo*

(a) 2D phantom image.

(b) 2D pre-resection image.

(c) 2D post-resection image.

Fig. 1. Example of 2D images used in this study.

Table 1. Phantom registration errors. Left, compression from the top; Right, compression from the top and the bottom. d_{in}, before registration; ϵ, after registration.

$Landmarks$	$d_{in}[pixels]$	$\epsilon[pixels]$	$d_{in}[pixels]$	$\epsilon[pixels]$
1	13.00	0.45	2.00	0.41
2	13.00	0.25	1.41	0.38
3	15.00	0.18	7.62	0.38
4	14.00	0.17	4.47	0.40
5	10.00	0.10	1.00	0.10
6	15.00	0.67	8.60	0.36
7	12.00	0.23	4.47	0.33
$Mean \pm SD$	13.14 ± 1.77	0.29 ± 0.20	4.23 ± 3.00	0.34 ± 0.11

cases each involve a pair of intra-operative images. In each case the first image has been acquired early in the procedure (a short time after craniotomy) and the second later in the procedure, typically after tumor resection. Figure 1(a) shows the phantom in its custom made clamp. Figure 1(b) and 1(c) show one of the *in vivo* cases. Panel 1(b) is the early image, panel 1(c) the later image. Note the large whole in the image shown on panel 1(c) that is the site of the resection.

3 Results

Qualitative and quantitative results obtained with the phantom and the *in vivo* images are presented in figures 2 and 3 and in tables 1 to 4. In figure 2, the left panels show the phantom in its original state. The right panels show the phantom after compression. The top row illustrates compression on one side, the bottom row compression on both side. Sulcal lines have been drawn on the compressed images and copied on the other ones. The middle panels show the results obtained when the images shown on the left panels are registered to the images shown on the right panels. Correct placement of the sulcal lines on the deformed images indicates a good registration between the deformed and

Table 2. Registration error for the first *in vivo* case; d_{in}, prior to registration; ϵ_r, after rigid body registration; ϵ_{nr} after non rigid registration.

$Landmarks$	$d_{in}[pixels]$	$\epsilon_r[pixels]$	$\epsilon_{nr}[pixels]$
1	16.13	6.83	0.38
2	33.54	6.93	0.22
3	19.31	7.15	0.25
4	14.21	8.51	0.34
5	17.46	9.99	0.50
6	25.55	5.02	0.54
7	36.77	0.11	0.30
$Mean \pm SD$	23.28 ± 8.90	6.36 ± 3.16	0.36 ± 0.12

Table 3. Registration error for the second *in vivo* case; d_{in}, prior to registration; ϵ_r, after rigid body registration; ϵ_{nr} after non rigid registration.

$Landmarks$	$d_{in}[pixels]$	$\epsilon_r[pixels]$	$\epsilon_{nr}[pixels]$
1	66.29	9.10	0.40
2	65.80	10.23	0.32
3	64.82	11.79	0.31
4	62.80	13.32	2.25
5	61.22	12.08	0.65
6	59.67	11.87	0.51
7	56.22	12.27	0.25
$Mean \pm SD$	62.40 ± 3.64	11.52 ± 1.40	0.67 ± 0.71

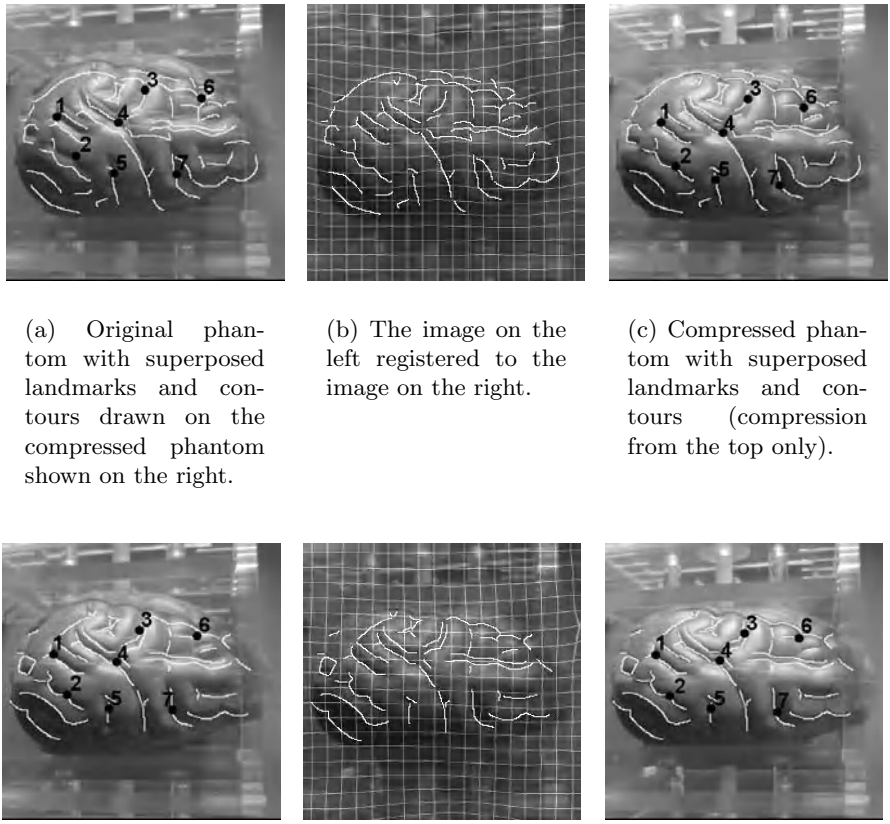
undeformed images. The middle images also show a regular grid defined on the undeformed image to which the deformation field is applied to show the smoothness and regularity of the transformation. Seven homologous landmarks (respectively x_i and y_i) also shown on the figure have been identified on the undeformed and deformed images for quantitative evaluation of the method we propose. These points have been chosen to be easily identifiable in the images and repeated selection of these landmarks resulted in a negligible localization error. Quantitative evaluation was performed as follows. The deformation field ($\mathcal{T}$) found using the previously described method was used to project the points x_i onto the deformed image to find the deformed points.

$$x_i' = \mathcal{T}(x_i) \tag{1}$$

The error for each pair of points (ϵ_i) is computed as the Euclidian distance between the manually selected points y_i on the deformed image and the corresponding transformed points x_i' as follows

$$\epsilon_i = \|(y_i - \mathcal{T}(x_i))\| \tag{2}$$

Table 1 shows the registration errors before (d_{in}) and after (ϵ) non-rigid registration. Figure 3 shows the results obtained for the three *in vivo* cases. The top panels show the images acquired early in the procedure and the bottom



(a) Original phantom with superposed landmarks and contours drawn on the compressed phantom shown on the right.

(b) The image on the left registered to the image on the right.

(c) Compressed phantom with superposed landmarks and contours (compression from the top only).

(d) Original phantom with superposed landmarks and contours drawn on the compressed phantom shown on the right.

(e) The image on the left registered to the image on the right.

(f) Compressed phantom with superposed landmarks and contours (compression from the top and the bottom).

Fig. 2. Registration results for the phantom case.

panels show the second images acquired later. The second row shows the results obtained after rigid-body registration of the images shown on the top to those shown on the bottom. The third row shows the final results obtained after both rigid and non-rigid registration. Sulcal lines have been drawn on the bottom images and copied on the other ones. This figure shows that serial intra-operative images can be very different from each other because of large resections (other factors include the appearance and/or disappearance of surgical instruments within the field of view or the application of clamps). This presents particular challenges to intensity-based registration algorithms as the one we use. In this work we found it necessary to outline manually Regions of Interest (ROIs) to

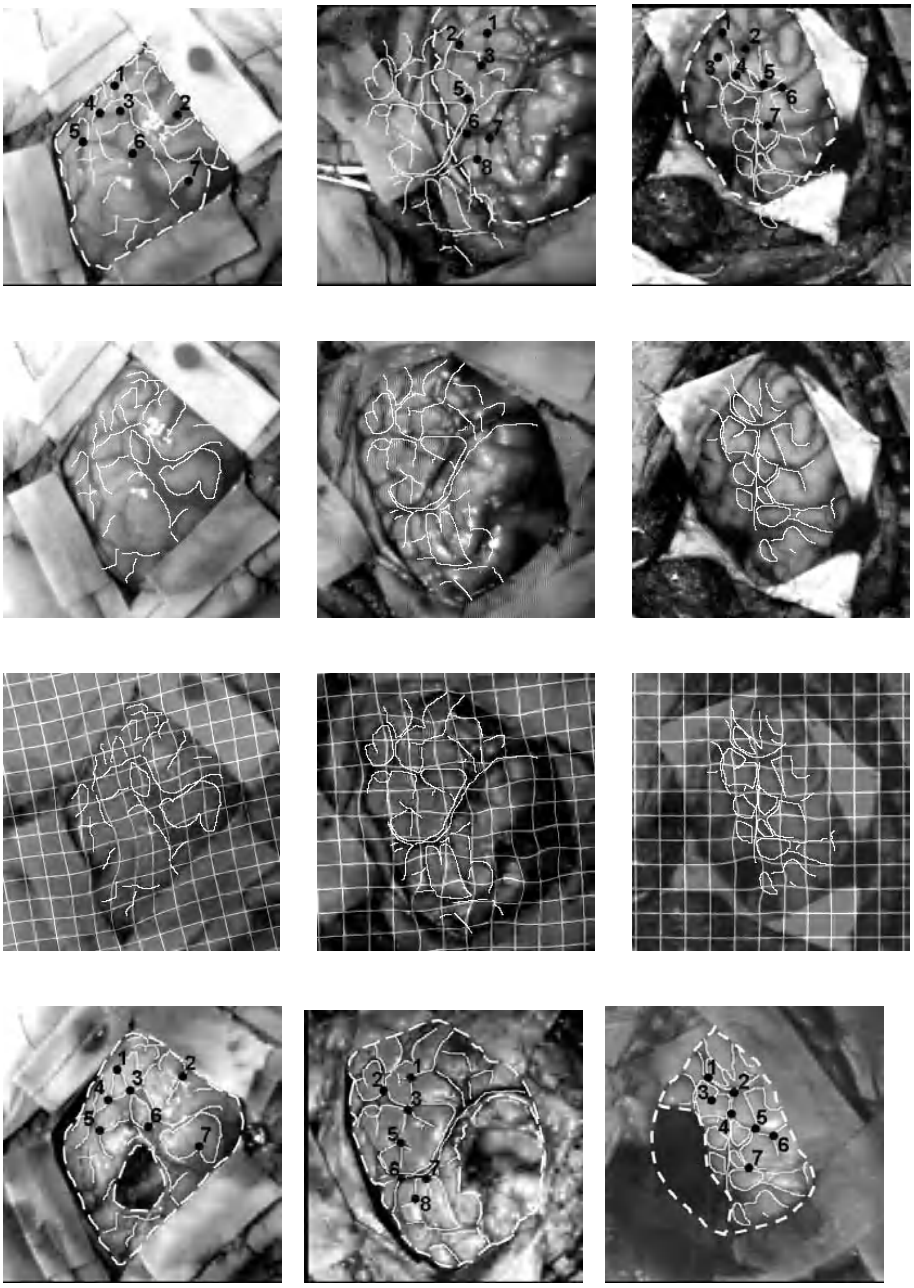


Fig. 3. *In vivo* undeformed and deformed images after rigid and non rigid registration. Landmarks used for validation have been superposed as well as corresponding contours.

Table 4. Registration error for the third *in vivo* case; d_{in}, prior to registration; ϵ_r, after rigid body registration; ϵ_{nr} after non rigid registration.

Landmarks	$d_{in}[pixels]$	$\epsilon_r[pixels]$	$\epsilon_{nr}[pixels]$
1	38.60	2.24	0.11
2	39.29	1.00	0.51
3	40.52	1.00	0.44
4	42.72	2.82	0.53
5	40.52	1.00	2.18
6	41.98	2.24	0.24
7	39.56	1.41	2.20
$Mean \pm SD$	40.46 ± 1.47	1.67 ± 0.75	0.89 ± 0.90

specify regions over which the transformations are computed. The dashed lines shown on the figure define these ROIs. Homologous landmarks have also been selected on the top and bottom images to permit quantitative evaluation of the registration results. Tables 2, 3, and 4 present the quantitative results for the *in vivo* data sets. In each of these tables, d_{in} refers to the registration error prior to registration, ϵ_r is the registration error after rigid body registration, and ϵ_{nr} is the registration error after both rigid and non-rigid registration. The large error prior to rigid body registration is due to the fact that the scanner was not placed at the same position for the first and second image acquisition.

4 Discussion

The results presented in this paper indicate that automatic intra-operative tracking of brain motion using a LRS is feasible. Despite large differences in the images due to resection and different viewing angles the approach we propose is robust enough to lead to sub-pixel registration errors. For the *in vivo* cases, the algorithm still requires manual intervention to delineate regions of interests over which the transformations are computed but these ROIs do not need to be delineated very carefully. Further development will address this issue. Also, the validation performed in this study is incomplete. It evaluates the quality of the 2D registration between the images but it does not quantify the overall 3D tracking error. In a companion paper [17] we evaluate this error for phantom data using an OPTOTRACK 3020 (Northern Digital Inc, www.ndigital.com) localization system and we show sub-millimetric differences between these measurements and the ones obtained with the method we propose. Intra-operative validation of the overall tracking accuracy of this method is ongoing. Should this study confirm our current results it would be an important step toward using LRS technology for real-time intra-operative brain tracking.

Acknowledgements

Supported in part by grants from the National Institutes of Health NIH/NCI 1R21 CA89657-01A2 and the Vanderbilt University Discovery Grant.

References

1. Kelly, P.J., Kall, B.A., Goerss, S.J., Earnest, F.I.: Computer-assisted stereotaxic laser resection of intra-axial brain neoplasms. Journal of Neurosurgery **64** (1986) 427–439
2. Nauta, H.J.: Error assessment during "image guided" and "imaging interactive" stereotactic surgery. Computerized Medical Imaging and Graphics **18** (1994) 279–287
3. Roberts, D.W., Hartov, A., Kennedy, F.E., Miga, M.I., Paulsen, K.D.: Intraoperative brain shift and deformation: a quantitative analysis of cortical displacement in 28 cases. Journal of Neurosurgery **32** (1998) 749–760
4. Nimsky, C., Ganslandt, O., Buchfelder, M., Fahlbusch, R.: Intraoperative magnetic resonance tomography - experiences in neurosurgery. Nervenarzt **71** (2000) 987–994
5. Nabavi, A., Mamisch, C.T., Gering, D.T., Kacher, D.F., Pergolizzi, R.S., Wells, W.M., Kikinis, R., Black, P.M., Jolesz, F.A.: Image-guided therapy and intraoperative mri in neurosurgery. Minimally Invasive Therapy and Allied Technologies **9** (2000) 277–286
6. Bucholz, R.D., Yeh, D.D., Trobaugh, J., McDurmont, L.L., Sturm, C.D., Baumann, C., Henderson, J.M., Levy, A., Kessman, P.: The correction of stereotactic inaccuracy caused by brain shift using an intraoperative ultrasound device. Lecture Notes in Computer Science: CVRMED-MRCAS '97 **1205** (1997) 459–466
7. Gobbi, D.G., Comeau, R.M., Peters, T.M.: Ultrasound/MRI overlay with image warping for neurosurgery. Lecture Notes in Computer Science: Proceedings of Medical Image Computing and Computer-Assisted Intervention: MICCAI '00 **1935** (2000) 106–114
8. Lunsford, L.D., Martinez, A.J.: Stereotactic exploration of the brain in the era of computed tomography. Surg Neurol **22** (1984) 222–230
9. Meyer, C.R., Boes, J.L., Kim, B., Bland, P.: Probabilistic brain atlas construction: Thin-plate spline warping via maximization of mutual information. Lecture Notes in Computer Science: Proceedings of Medical Image Computing and Computer-Assisted Intervention: MICCAI '99 **1679** (1999) 631–637
10. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Nonrigid registration using free-form deformations: Application to breast mr images. IEEE Transactions on Medical Imaging **18** (1999) 712–721
11. Miga, M.I., Paulsen, K.D., Kennedy, F.E., Hoopes, P.J., Hartov, A., Roberts, D.W.: Initial in-vivo analysis of 3d heterogeneous brain computations for model-updated image-guided neurosurgery. Lecture Notes in Computer Science: Proceedings of Medical Image Computing and Computer-Assisted Intervention: MICCAI '98 **1496** (1998) 743–752
12. Ferrant, M., Nabavi, A., Macq, B., Black, P.M., Jolesz, F.A., Kikinis, R., Warfield, S.K.: Serial registration of intraoperative mr images of the brain. Medical Image Analysis **6** (2002) 337–359

13. Galloway, R.L., Macuinas, R.J., Bass, W.A., Carpini, W.: Optical localization for interactive, image-guided neurosurgery. Medical Imaging **2164** (1994) 137–145
14. Miga, M.I., Sinha, T.K., Cash, D.M., Galloway, R.L., Weil, R.J.: Cortical surface registration for image-guided neurosurgery using laser range scanning. IEEE Transactions on Medical Imaging (2003) In press.
15. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality image registration by maximization of mutual information. IEEE Transactions on Medical Imaging **16** (1997) 187–198
16. Rohde, G.K., Akram, A., Dawant, B.M.: The adaptive bases algorithm for intensity based nonrigid image registration. IEEE Transactions on Medical Imaging (2003) In press.
17. Sinha, T.K., Duay, V., Dawant, B.M., Miga, M.I.: Cortical shift tracking using a laser range scanner and deformable registration methods. Lecture Notes in Computer Science: Proceedings of Medical Imaging Computing and Computer Assisted Intervention: MICCAI '03 (2003) Submitted.

Inverse Consistent Image Registration of MR Brain Scans: Handedness in Normal Adult Males

Xiujuan Geng[1], Dinesh Kumar[1], Gary E. Christensen[1], and Michael W. Vannier[2]

[1] Department of Electrical and Computer Engineering
The University of Iowa, Iowa City, IA, 52242
{xiujuan-geng,dinesh-kumar,gary-christensen}@uiowa.edu
[2] Department of Radiology
The University of Iowa, Iowa City, IA, 52242
michael-vannier@uiowa.edu

Abstract. Brain shape differences between right and left-handed normal adults were evaluated by inverse-consistent linear-elastic image registration (ICLEIR) applied to MRI scans from two groups. The study populations were 9 right-handed and 9 left-handed adult males from ages of 24 to 51 years old. The mean brain shape of each population was computed and used as the reference shape for detecting shape differences. Nonrigid, ICLEIR transformations that registered the mean brain image with the brain images from the pooled populations were used to detect local brain population shape differences. Following the approach of Thirion et al., asymmetry maps between the left and right hemispheres of a brain image were computed by registering each brain image with their mirror images. Local statistical shape differences between the two populations were determined using one and two-tailed t-tests at each voxel in the coordinate system of the mean brain shape. Four t-tests were computed and compared which included the log-Jacobian and magnitude-divergence of the individual-to-pooled-average (IPA) correspondence map and the log-Jacobian and magnitude-divergence of the asymmetry maps. Local shape differences between populations were evaluated to determine the location of asymmetries due to handedness. Statistically significant ($\alpha = 0.01$) shape differences were found in this small pilot study with a sample size of 9 for each group. Although the populations were too small to draw conclusions regarding neuromorphological differences between left and right handed individuals, the method shows promise for detecting brain shape differences between different populations.

1 Introduction

It is important to detect subtle shape differences between populations to understand neuromorphological shape and its variability. This paper describes a statistical method for detecting significant shape differences between populations of brain images. Hypothesis tests performed indirectly on transformation fields are used to detect shape differences between the populations.

Many methods are used to study anatomical shape differences in medical images. The most straight forward way to measure neuroanatomical shape differences is by manual measurement [1]. In addition, methods are available that analyze the regional volume

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 71–80, 2003.

of gray and white matter densities in stereotaxic space after spatial normalization. These methods include regional analysis of volumes examined in normalized space (RAVENS) [2] and Voxel-based morphometry (VBM) [3] These methods are less sensitive to errors during spatial normalization, but they are not able to detect higher order shape changes. Still other methods detect quantitative shape differences by analyzing transformations [4,5,6,7] that register a template to a target image. These methods provide better localization of shape differences but are sensitive to the transformation accuracy.

The method presented in this work follows the seminal work of Thirion et al. [4,5]. In their approach, shape and asymmetry differences are analyzed by performing statistical t-tests on the magnitude-divergence of asymmetry maps. In this work, statistical shape differences are detected using T-statistic tests applied to the log-Jacobian and magnitude divergence of asymmetry maps and to individual-to-pooled-average (IPA) correspondence maps. Inverse consistent linear-elastic image registration [8] is used to estimate the IPA correspondence maps.

The rest of the paper is organized as follows. Section 2 provides an overview of our method for detecting statistically significant shape differences between two populations of images based on inverse-consistent linear-elastic image registration (ICLEIR) correspondence maps. Experimental results are presented in Section 3 that detect statistically significant shape differences between 3D MRI brain images collected from left and right handed human males between the ages of 24 and 51. Finally, the paper is summarized and conclusions are drawn in Section 4.

2 Methods

This section describes our method for detecting statistically significant shape differences between populations of brain images using image registration and statistical hypothesis testing. A brief outline of the method is given below followed by a more detailed description of each step.

1. Synthesize a pooled average brain image representing the mean brain shape of all the brains from both populations. The pooled average brain image is produced by averaging inverse-consistent linear-elastic image registration (ICLEIR) transformations between a template image and the images from the pooled population.
2. Estimate the ICLEIR IPA correspondence maps that register the pooled average image with each image in the pooled population.
3. Compute the asymmetry maps by transforming each data set into its mirror image using ICLEIR registration.
4. Compute the log-Jacobian and magnitude-divergence images for each IPA correspondence map and asymmetry map in the coordinate system of the pooled average brain image.
5. Detect statistically significant areas of shape differences by computing the two population univariate t-test at each voxel in the coordinate system of the pooled average brain image. T-tests are performed on the log-Jacobian and magnitude divergence of the IPA correspondence maps and asymmetry maps.

2.1 Synthesizing the Pooled Average Brain Image and IPA Correspondence Maps

Inverse-consistent linear-elastic registration [8] is used to calculate transformations between template and target images. This registration method jointly estimates the forward transformation and the reverse transformation while minimizing the inverse consistency error. Figure 1 shows an example transformation result registering a template image to a target image. The deformed image should be identical to target image, however the intensity difference image between them is not equal to zero everywhere.

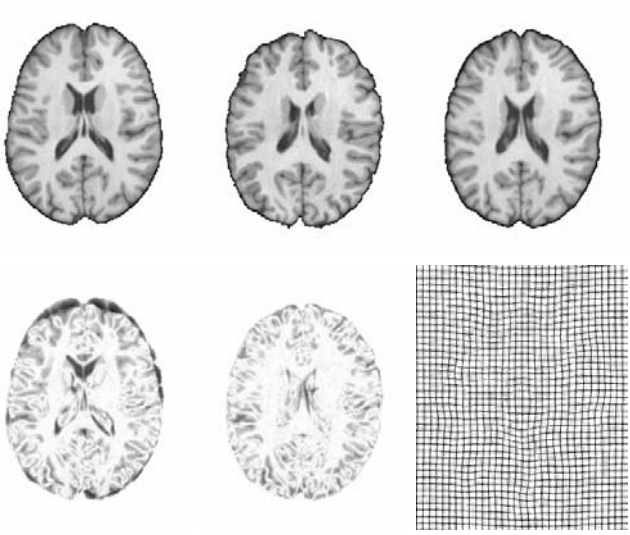


Fig. 1. Transformation from template image to target image. In the top row, from left to right: template image T_i, target image T_j, and deformed template image $T_i(h_{i,j})$. In the bottom row, from left to right: magnitude intensity difference between T_i and T_j, intensity difference between $T_i(h_{i,j})$ and T_j, and deformed grid image.

The pooled average brain image of the pooled population is computed from the transformations between a template image and each image in the pooled population following the algorithm in [9]. A template image T_i from one of the populations is registered with each image T_j in the pooled population producing a set of Lagrangian transformations $g_{ij}(x)$ from the coordinate system of image T_i to that of image T_j. These transformations are averaged together producing an average transformation using the equation $g_{i,\bar{i}}(x) = \frac{1}{N}\sum_{j=1}^{N} g_{ij}(x)$. The template image T_i is transformed into the average shape $\overline{T}_i$ of the pooled population using the equation $\overline{T}_i(x) = T_i(h_{i,\bar{i}}(x))$ where the Eulerian transformation $h_{i,\bar{i}}(x)$ is equal to the inverse of the Lagrangian transformation $g_{i,\bar{i}}(x)$.

The pooled average brain image was registered with each data set in the combined population using ICLEIR algorithm producing high-dimensional, nonrigid, individual-to-pooled-average (IPA) correspondence maps.

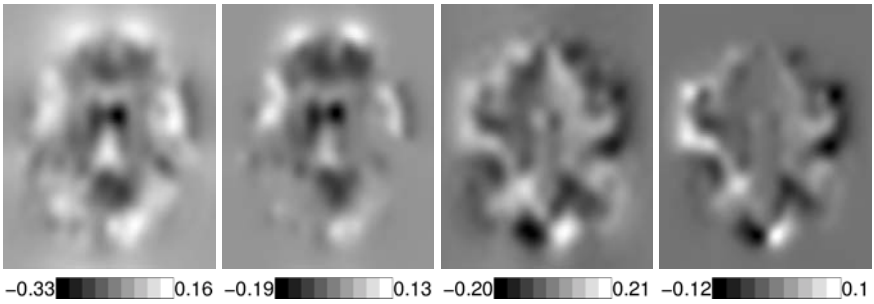


Fig. 2. Log Jacobian and magnitude-divergence images of the IPA correspondence map and asymmetry map of one of the right handed individuals. From the left to right: log-Jacobian image for a transformation between two images, magnitude divergence image for the same transformation , log-Jacobian image for a asymmetry map of a image, magnitude-divergence image for the same asymmetry map. Note: all the values in the color bars for the two magnitude-divergence images are multiplied by 10^4.

2.2 Asymmetry Maps

We follow the approach of Thirion et al. [4,5] to compute the dissymmetry or asymmetry map for each image in the pooled population. The asymmetry map is computed by registering each image with its mirror image[1] with respect to the midsagittal plane using ICLEIR registration [8]. The ICLEIR transformation between the image and its mirror image is defined as the asymmetry map of the image.

Figure 2 shows the log-Jacobian and magnitude-divergence images of the IPA correspondence map and the asymmetry map for a data set from population 1. The log-Jacobian operator and the magnitude-divergence operator applied to the asymmetry map measure the asymmetry between the left and right brain hemispheres. The larger the magnitude value is at a point in the image, the more asymmetrical the left and right hemispheres are at that point. The left and right sides of the brain are assumed to be symmetric at a point if the log-Jacobian and/or the magnitude-divergence of the asymmetry map at that point is zero. Notice that the log-Jacobian and magnitude divergence images of the asymmetry map are symmetric about the midsagittal plane. The symmetry about the midsagittal plane provides a qualitative validation that the left hemisphere is properly mapped into the shape of the right hemisphere and vice versa. Also, notice that the log-Jacobian and magnitude operators produce images that are visually consistent with each other.

2.3 Log-Jacobian and Magnitude-Divergence Operators

The log-Jacobian and magnitude-divergence operators can be used to describe local shape change present in high-dimensional nonrigid transformation (see Fig. 1). Examining these operators reveals their similarities and differences. For brevity of presentation, the Jacobian and magnitude-divergence operators are described in 2D. Let

[1] The mirror image was computed by flipping the image in the x-voxel dimension and shifting the flipped image by one voxel in the x-voxel dimension.

$h(x,y) = (x + u_x(x,y), y + u_y(x,y))$ denote a Eulerian transformation and $u(x,y) = (u_x(x,y), u_y(x,y))$ denote its displacement field. The Jacobian of the transformation h is defined as

$$J(h) = \det \begin{bmatrix} 1 + \frac{\partial u_x}{\partial x} & \frac{\partial u_x}{\partial y} \\ \frac{\partial u_y}{\partial x} & 1 + \frac{\partial u_y}{\partial y} \end{bmatrix} = 1 + \frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} + \frac{\partial u_x}{\partial x}\frac{\partial u_y}{\partial y} - \frac{\partial u_x}{\partial y}\frac{\partial u_y}{\partial x}. \quad (1)$$

The Jacobian describes the local expansion and contraction that occurs when transforming the template into the shape of a target image. The log of the Jacobian is computed to linearize the differential volume change described by the Jacobian operator. For Eulerian transformations, log-Jacobian values less than zero correspond to expansion and values greater than zero correspond to contraction.

The magnitude-divergence operator first presented in [4,5] weighs the divergence operator more in regions with large displacements compared to regions with small displacements. The idea behind this weighting scheme is to amplify regions in the transformation where there is large shape changes and attenuate regions with small shape changes. The expression for the magnitude-divergence operator is given by

$$||u||\nabla \cdot u = \sqrt{u_x^2 + u_y^2}(\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y}). \quad (2)$$

Notice that Eqs. 1 and 2 are similar in that both have the terms $\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y}$. However, the magnitude-divergence operator does not have the product derivative terms present in Eqs. 1. This comparison demonstrates that although the magnitude-divergence operator is able to detect certain regions of shape change, its physical interpretation is not as straight forward as the log-Jacobian which detects changes based on volumetric differences. The results presented in this paper also suggest that log-Jacobian may be more sensitive to shape differences than the magnitude-divergence operator due to these additional product derivative terms.

2.4 Hypothesis Testing

Univariate t-tests are performed on the log-Jacobian and magnitude-divergence operators of the IPA and asymmetry maps to detect statistically significant regions of shape differences. At each voxel in the coordinate system of the pooled average image, set the values of the operator to be $(x_1^j, ..., x_{n_j}^j)$ for individuals of population $j = 1, 2$. We assume that $x_1^j, ..., x_{n_j}^j$ are i.i.d. normally distributed with unknown mean μ_j and variance σ_j^2. The T-statistic is computed as:

$$T = \frac{(\bar{X_1} - \bar{X_2}) - (\mu_1 - \mu_2)}{\sqrt{({S_1}^2/n_1 + {S_2}^2/n_2)}} \quad (3)$$

where $\bar{X}_j = \sum_{i=1}^{n_j} X_i^j$ are the sample means and $S_j^2 = \frac{1}{n_j - 1}\sum_{i=1}^{n_j}(X_i^j - \bar{X}_j)^2$ are the sample variances of population $j = 1, 2$. The T-statistic has a t-distribution with $\nu = \frac{({s_1}^2/n_1 + {s_2}^2/n_2)^2}{[({s_1}^2/{n_1}^2/(n_1-1))] + [({s_2}^2/{n_2}^2/(n_2-1))]}$ degrees of freedom.

The null hypothesis $H_0 : \mu_1 = \mu_2$ assumes no difference between the populations and the alternative hypothesis $H_1 : \mu_1 \neq \mu_2$ assumes there is a difference. Under these hypotheses, $\mu_1 - \mu_2 = 0$ in Eq. 3. The alternative hypothesis is selected if the probability of the observed T-statistic value is less than or equal to α under the T-distribution.

3 Results

To distinguish the effects of handedness on adult male brain structure, we randomly selected 18 T1-weighted 3D MR images of the normal male brains (age range: 24-51): 9 right-handed and 9 left-handed (see Fig. 3). These data sets were taken from the Neurology Department at the University of Iowa. Due to the small sample size, the results in this section illustrate the potential of our algorithm rather than give medically significant results.

Fig. 3. Transverse images taken from 4 of the 18, 3D MRI data sets analyzed in this paper. Columns 1 and 2 show data from two right-handed males (from population 1) and columns 3 and 4 show data from two left-handed males (from population 2).

All the 3D MRI data sets were preprocessed by resizing them to a $256 \times 320 \times 256$ voxel volume with $0.7mm \times 0.7mm \times 0.7mm$ voxel dimensions. The data sets were rigidly rotated and translated to orient the brain to the voxel lattice using the anterior commissure(AC), posterior commissure(PC), and the inter-hemispheric fissure. Next, the data sets were translated to align the AC point with coordinate (128,184,128).

3.1 Single Population Asymmetry

Our first experiment studied the asymmetries of left and right-handed populations separately. Each data set in the populations was registered to their mirror image using inverse-consistent linear-elastic image registration (ICLEIR) producing asymmetry maps. Next, a pooled average brain image was constructed that represented the average shape of 18 left and right-handed brain populations using the procedure discussed above. Individual to pooled average (IPA) correspondence maps were computed using ICLEIR registration and were used to map the log-Jacobian and the magnitude-divergence asymmetry maps to the pooled average brain image coordinate system.

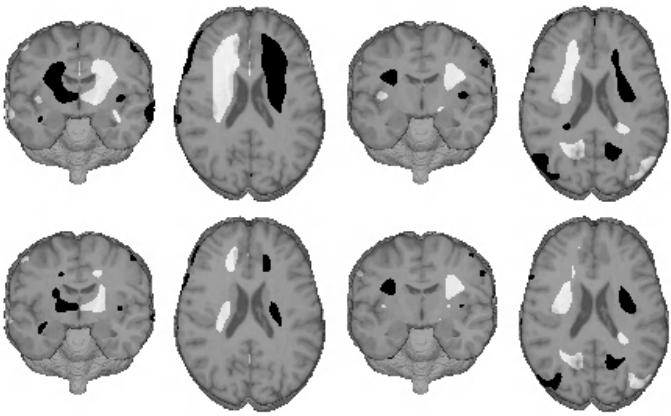


Fig. 4. Regions of significant brain asymmetry for populations of 9 right-handed males (columns 1 and 2) and a population of 9 left-handed males (columns 3 and 4). Regions of white and black correspond to significant ($\alpha = 0.005$) expansion and contraction, respectively, between the left and right brain hemispheres. The top row was produced using log-Jacobian asymmetry maps and the bottom row using magnitude-divergence asymmetry maps.

Figure 4 shows the single population t-tests that were performed at each voxel on the log-Jacobian and magnitude-divergence images for the left and right-handed populations. Regions of white correspond to a one-tailed t-test in which there was a significant ($\alpha = 0.005$) expansion between the left and right hemispheres. Similarly, regions of black correspond to a one-tailed t-test in which there was a significant ($\alpha = 0.005$) contraction between the left and right hemispheres. The fact that the asymmetry maps are nearly symmetric about the midsagittal plane serves as a qualitative validation of the method. That is, if region 1 is significantly larger than its corresponding region 2 in the opposite hemisphere, then region 2 should be be significantly smaller than the original region 1.

Notice that the regions of significant asymmetries between the right and left hemispheres in both left and right-handed populations are similar. However, the right-handed brains have larger asymmetrical regions than the corresponding regions for the left-handed populations. In addition, it appears that the asymmetries are more spread out in left-handed brains than in right handed ones. Comparing the results of the two operators, we noticed that the log-Jacobian operator always gave a larger response than the magnitude-divergence operator.

3.2 Left vs. Right-Handed Population Comparison

Significant shape differences between the left and right-handed brain populations were investigated using a two population t-test. Four univariate t-tests were computed at each voxel in the coordinate system of the pooled average brain image. These t-tests were computed on the (1) log-Jacobian of the asymmetry maps, (2) magnitude-divergence of asymmetry maps, (3) log-Jacobian of the IPA correspondence maps, and (4) magnitude-

Fig. 5. Regions of significant asymmetry between populations of 9 left-handed and 9 right-handed males. White shows regions where the left-handed population has a significantly ($\alpha = 0.005$) larger asymmetry map value than right-handed population and black shows regions where the right-handed population has significantly ($\alpha = 0.005$) larger asymmetry map value than the left-handed population. Panels 1 and 2 show the statistical response produced using the log-Jacobian asymmetry maps and panels 3 and 4 show the response using the magnitude-divergence asymmetry maps.

Fig. 6. Regions of significant brain volume difference determined by the log-Jacobian operator (panels 1 and 2) and magnitude-divergence differences (panels 3 and 4) between populations of 9 left and 9 right-handed males. White corresponds to brain regions of significant ($\alpha = 0.005$) expansion of the left-handed population compared to the right-handed population. There were no regions of significant ($\alpha = 0.005$) contraction of the left-handed population compared to the right-handed population.

divergence of the IPA correspondence maps. The results of these t-tests are shown in Figs. 5 and 6.

Figure 5 shows regions of significantly different asymmetries between the left and right-handed populations in white and black. Note, that it is not possible to determine whether the right- or left-handed population is more asymmetric in a white or black region. The white and black regions are nearly symmetric about the midsagittal plane due to the left/right symmetry of the asymmetry maps. Notice that both log-Jacobian or magnitude-divergence operators give similar results. However, the response of the log-Jacobian operator seems to be slightly larger than that of the magnitude-divergence operator.

Compared to the results of one population statistical tests, the two population tests have much less response under the same significant level. One reason for this is because both populations have the same asymmetry trend, i.e., most white regions are in the left brain hemisphere for both the left and right-handed populations (see Fig. 4).

Figure 6 shows regions of significant shape differences between the left and right-handed populations as determined by computing the two population t-test on the IPA cor-

respondence maps. White corresponds to brain regions of significant ($\alpha = 0.005$) expansion of the left-handed population compared to the right-handed population. There were no regions of significant ($\alpha = 0.005$) contraction of the left-handed population compared to the right-handed population suggesting that the average right-handed brain is smaller than the average left-handed brain in several regions. This result is probably due to the fact that the data sets were not scaled to the same dimension in this study. Also note that, the slices selected to show significant differences between the IPA correspondence map are the ones which have the largest responses and do not correspond to the slices shown for the asymmetry map results.

These results show that a significant shape difference in the left hemisphere does not necessarily imply a corresponding shape difference in the right hemisphere. Therefore, the two population t-tests on the asymmetry map provide different information than the two population t-tests on the IPA correspondence maps. The former t-test detects differences in hemispheric symmetry between the populations while the latter test detects shape differences between the populations. The former test produces a much larger response than the latter, which implies there are more shape differences between the left and right hemispheres within the same brain than between left (right) hemispheres of different brains.

4 Summary and Conclusions

We presented several methods to detect asymmetry and volume differences between populations of brain images of left and right handed individuals using statistical t-distribution tests. We compared the magnitude-divergence operator method of Thirion et al. to the log-Jacobian operator. Inverse-consistent linear-elastic image registration was used to (1) construct a pooled average brain image of the combined left and right handed populations, (2) compute asymmetry maps between the left and right brain hemispheres within each individual, and (3) determine an individual to pooled average (IPA) correspondence map. Statistical tests were computed using the log-Jacobian and magnitude divergence of the asymmetry map and the IPA correspondence map. The log-Jacobian operator studied in this work characterizes the local volume differences between the left and right hemispheres within an individual brain and between the individual and the pooled population average brain.

Although our population sizes were small (9 left and 9 right handed individuals), the method was able to detect statistically significant ($\alpha = 0.01$) regions of shape differences between the populations. This statistical result was not obvious from visual inspection since the asymmetry maps of left and right handed individuals visually appeared very similar. The log-Jacobian and magnitude-divergence operators detected similar statistically significant regions of shape difference and the log-Jacobian operator seemed to provide a larger area of response compared to the magnitude divergence operator.

The methodology presented in this paper is general and can be applied to detect morphological shape differences between populations. Example applications include detecting differences between normal and abnormal populations, shape differences that occur over time within and between individuals, and shape changes due to medical treatment.

Acknowledgments

We would like to thank Russ Lenth, Director of the Statistical Consulting Center, The University of Iowa for his helpful comments. This work was supported in part by the NIH grants NS35368, DC03590, and HL64368. We would like to thank Richard A. Robb of the Mayo Clinic, Rochester, MN for his continued collaboration in providing the Analyze^{TM} software package.

References

1. Allen JS, Damasio H, and Grabowski TJ, "Normal neuroanatomical variation in the human brain: an mri-volumetric study," *Am J Phys Anthropol*, vol. 118, pp. 341–58, 2002.
2. Christos Davatzikos, Ahmet Genc, Dongrong Xu, and Susan M. Resnick, "Voxel-based morphometry using the ravens maps: Methods and validation using simulated longitudinal atrophy," *NeuroImage*, vol. 14, pp. 1361–1369, 2001.
3. J. Ashburner and K.J. Friston, "Voxel-based morphometry - the methods," *NeuroImage*, vol. 11, no. 6, pp. 805–821, 2000.
4. Jean-Philippe Thirion and Neil Roberts, "Statistical analysis of normal and abnormal dissymmetry in volumetric medical images," *Medical Image Analysis*, vol. 4, pp. 111–121, 2000.
5. Sylvain Prima, Jean-Philippe Thirion, Gerard Subsol, and Neil Roberts, "Automatic analysis of normal brain dissymmetry of males and females in mr images," in *MICCAI'98, LNCS 1496*, William M. Wells, Alan C. F. Colchester, and Scott Delp, Eds., Berlin Heidelberg, 1998, pp. 770–779, Springer-Verlag.
6. P.M. Thompson and A.W. Toga, "Detection, visualization and animation of abnormal anatomic structure with a deformable probabilistic brain atlas based on random vector field transformations," *Medical Image Analysis*, vol. 1, no. 4, pp. 271–294, 1997.
7. J.W. Haller, A. Banerjee, G.E. Christensen, M. Gado, S.C. Joshi, M.I. Miller, Y. Sheline, M.W. Vannier, and J.G. Csernansky, "3D hippocampal morphometry by high dimensional transformation of a neuroanatomical atlas," *Radiology*, vol. 202, no. 2, pp. 504–510, Feb. 1997.
8. G.E. Christensen and H.J. Johnson, "Consistent image registration," *IEEE Transactions on Medical Imaging*, vol. 20, no. 7, pp. 568–582, July 2001.
9. G.E. Christensen and H.J. Johnson, "Synthesizing average 3D anatomical shapes," *Submitted. IEEE Transactions on Medical Imaging*.

Simultaneous Population Based Image Alignment for Template Free Spatial Normalisation of Brain Anatomy

C. Studholme

Dept. Radiology, University of California, San Francisco

Abstract. Current approaches to spatial normalisation of brain images have made use of a target image to which each subject image is matched. However, in many cases the use of a single brain template, or a statistical one derived from multiple subjects of another population, does not adequately capture the structure present in a population of anatomies under investigation. In such cases this paper proposes that a better approach may be to seek a method of driving subjects in the group into registration with each other, rather than with an unrepresentative template. This paper explores the approach of extending registration concepts from multimodality registration, specifically those deriving criteria from the joint probability distribution of image values, to the general case of describing the alignment of a population of images simultaneously. Geometric constraints forcing the convergence to an average geometric shape are discussed and results presented on synthetic images and clinical brain image data.

1 Introduction

Spatial normalisation of brain anatomy from different subjects [19,2,4,16,5,6] is an increasingly important step in automated brain image analysis [17,7], useful both in structural and functional data analysis. One important and challenging application of spatial normalisation, which is the focus of this work, is in computational morphometry of tissue loss in neurodegeneration [13]. In such studies, it is common to make use of a reference anatomy or shape, to which all subject anatomies are aligned in order to permit local mapping of shape difference across a population. However, in an aging population, a range of cumulative-random degenerative processes have acted on each anatomy, and a separate reference brain may have little in common with many subject anatomies being studied. The aim of this paper therefore is to explore an approach to spatially aligning structural MRI scans that is free from the use of a single reference shape (be it either a statistical combination of anatomies or a single example). Specifically, we explore a generic statistical approach to defining a measure of population alignment, by examining the joint probability distribution between all image values across a population. This is combined with geometric constraints to allow simultaneous spatial alignment of images of populations of randomly varying shapes.

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 81–90, 2003.

2 Method

2.1 Why Population Focused Alignment?

In applying deformation morphometry to study brain shape, particularly in dementia and aging, there are two key problems posed by normal aging and abnormal neurodegenerative conditions:

Varying Patterns of Regional Tissue Contrast: The presence of local white matter tissue contrast changes due, for example, to local vascular disease, or other white matter changes, cannot easily be captured by a single target anatomy. It may possible to construct a statistical target from another population [15], which somehow captures the local presence of regional white matter lesions. However, the question then arises how to align the individuals from this reference population in order construct this statistical target? More importantly, this statistical reference target will be derived from a different population to that of the disease being studied, in terms of lesions occurrence, location and contrast, and may therefore still be far from optimal for the purposes of study alignment.

Extreme Geometric Differences: Accurate estimation of mappings between boundaries in a reference brain to a subject is more difficult for subjects with greater shape variations arising from structural degeneration, than for more normal subject brains, simply because of the greater difference in curvature and separation of anatomical boundaries. The use of a synthetic average of a population can provide some reduction in the overall distance for the registration process to 'cross', by providing a more representative starting shape. However the distance between an average and a highly diseased individual may still be significant, and it is often the case that such individuals that are shape 'outliers' are important to capturing the nature of the shape variation in a population.

Fig. 1. Slices through a T1 weighted MR image of the brain of a cognitively normal 87 year old showing shape variations (in particular ventricular shape) and contrast variations (peri-ventricular white matter lesions, visible as a reduction in white matter intensity around the ventricles).

These two observations lead us to propose that a better solution may be one which attempts to simultaneously co-align subjects together in the population of interest, rather than with some other target shape. Such an approach may make

use of the fact that the population itself provides multiple examples of intermediate shapes, from average to extreme, between which local mappings can be more easily formed, as the processes of bringing all individuals together progresses. We can additionally make best use of any shared regional contrast properties within the actual population of interest to derive meaningful correspondences between these structures. This is particularly important in the context of morphometry and shape analysis, where it may reveal common shape properties in lesions which are themselves meaningless with respect to a separate target anatomy from that being studied. A common example of features of varying contrast and shape are per-ventricular white matter lesions which can occur in white matter, around its boundary with ventricular CSF. Such lesions are common in aging and dementia populations and ideally we would like to bring these diffuse regions in different individuals into correspondence to compare their size and extent around the ventricular system.

2.2 Population Focused Image Intensity Alignment

Given a set of images of different subjects, each described by an intensity value $i_n(\mathbf{x}_n)$, varying across the imaged space $\mathbf{x}_n \in X_n$ of a subject $n \in 1 \dots N$, we can define a common space of points $\mathbf{x}_R \in X_R$, and a set of transformations $\mathcal{T}$, from this space, to each of the image spaces

$$\mathcal{T}(\mathbf{x}_R) = \{T_{Rn} : \mathbf{x}_n = T_{Rn}(\mathbf{x}_R), n \in 1 \dots N\}, \tag{1}$$

so that the intensity of subject image n at location $\mathbf{x}_R$ is given by,

$$i'_n(\mathbf{x}_R) = i_n(T_{Rn}(\mathbf{x}_R)). \tag{2}$$

Each location can then be assigned a vector of length N values,

$$\mathbf{I}(\mathbf{x}_R) = \{i'_n(\mathbf{x}_R), n \in 1 \dots N\} \tag{3}$$

with one component derived from each subject image.

In this work we will examine the general case where these images are not from identical modalities, and therefore we cannot assume that the same tissue class exhibits the same intensity in each image. This is typically the case even for the same type of MRI sequence, where differences in gain and shim settings mean that intensity values cannot be calibrated and mapped to a standard physical unit. It is even more of an issue when relating MR images of subjects with varying tissue integrity. In order to derive a meaningful relationship between these intensities we take the approach explored in general multi-modality alignment, that of examining relationships between the image values in terms of their observed co-occurrence, as captured by the properties of the joint probability distribution of their intensity values [8]. In our case though, rather than being a simple 2 dimensional distribution, it is a distribution with dimension equal to the number of subjects within the population being studied such that:

$$p(\mathbf{I}) = p(i'_1, i'_2, \dots i'_n) = P(I_1 = i'_1, I_2 = i'_2, \dots I_n = i'_n) \tag{4}$$

For the purposes of considering the structure of this distribution in the following section we will assume a discrete set of tissue-intensity types.

Response of Histograms to Misalignment: An important observation arises when considering the properties of this joint distribution of intensities, with respect to the alignment and misalignment of the entire population of images. This can be seen by first considering the case of image pair alignment, as illustrated in the top of figure 2(a) for a simple two level, noise free image pair. As the pair of images are mis-aligned, regions of uniform intensity in each image do not fully overlap, but instead overlap with neighboring regions in the other image. This reduces the height of the primary peaks in the distribution and creates confounding peaks in the distribution, which correspond to the misalignment of intensities into neighboring regions. The height of both the primary peaks, p_{11} and p_{22}, and the confounding peaks, p_{12} and p_{21}, are then proportional to the mis-registration Δr. In fact, given no noise and simple displacement, the increase in the occurrence of confounding intensity pairs $(j, k), j \neq k$ is simply

$$\Delta p_{jk} = \frac{\Delta r}{2}, j \neq k. \tag{5}$$

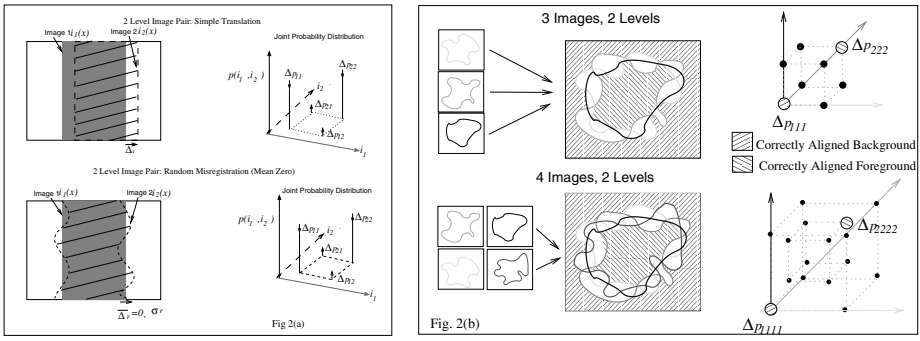


Fig. 2. (a) An illustration of the influence of simple (top) and random (bottom) misalignment of image boundaries on the joint intensity histogram, for 2 level images with no noise. (b) Peaks (in this case circles of varying size) in 3 and 4 dimensional histograms corresponding to correct and incorrect alignment of sets of 3 and 4 images. As the number of images increases, the mis-aligned fraction of image values is distributed over a larger number of combinations of values (at the corners of a hyper-cube).

Methods such as mutual information or other joint distribution derived criteria, rely on a starting estimate that ensures that the largest, most probable intensity classes in the two images that should correspond at registration, overlap more than any other tissue-intensity pair at the start of the registration. If this is not the case then there is no stable common structure to drive the images further into alignment. To ensure this, the size of the 'confounding' peaks in the histogram at the starting estimate must be lower than the main peak of correctly co-occurring intensities. The entropy criteria then drives the registration

in a direction which enhances the main peak by choosing transformations which reduce these confounding peaks.

For a given a random perturbation of the transformation between the images away from correct alignment, as illustrated in the bottom of figure 2(a): If the mean displacement of the boundaries $\bar{\Delta r} = 0$, but there is some variation around correct alignment along this boundary resulting in ρ_r of the values being mapped to the wrong tissue-intensity class in the other image, then the increase in the height of confounding peaks (the amount of space occupied by incorrectly overlapping intensities) is simply:

$$\Delta p_{jk} \propto \frac{\Delta \rho_r}{2}, j \neq k. \tag{6}$$

For the general case of a population of images, with a random shape variation, from which is created a multi-dimensional joint distribution of intensity values, the influence of the amount of misalignment can be examined by generalizing equation 6, as illustrated by Figure 2 (b). We see that, for a simple binary image boundary, there will be a total 2^N possible intensity sets for a given pixel or voxel. Two combinations can occur when the two values either side of a boundary are correctly aligned, and the remaining $2^N - 2$ will occur when different parts of the boundary in each image are mis-aligned, creating different combinations of intensity values. The general case of the size of these confounding peaks, p', created by misalignment of 2 intensity levels in N randomly deformed images can then be written images as:

$$\Delta p' \propto \frac{\Delta \rho_r}{2^N - 2}. \tag{7}$$

From this we can see that for any reasonably sized imaging study the denominator $(2^N - 2)$ to be large and the relative size of confounding peaks created in the histogram by misalignment will be small. Any remaining peaks in the joint distribution will represent the largest commonly overlapping intensity classes across the entire population of images at the starting estimate of the registration.

2.3 A Measure of Group Intensity Alignment

For a relatively large set of images, equation 7 tells us that, given a rough starting estimate that has the bulk of white matter and the head of each subject approximately aligned, then the set of co-occurring intensities at points in the mis-aligned folds of white and gray matter around the cortex will have very low probability of occurrence. The histogram, at the start of registration, will be dominated by peaks corresponding to deep white matter and air which will be commonly aligned across the entire population. To refine the collective alignment of tissues from this point we require a local deformation force which maps a reference location to a subject image, such that the overall probability of occurrence of the set of intensities at that point is increased with respect to the currently observed distribution of intensity sets. This work investigates a two

step iterative process to achieving this, given a starting set of transformations $\mathcal{T}(\mathbf{x_R})$: Estimation of the joint intensity distribution followed by a local deformation step toward a less dispersed joint histogram using measures derived from the current distribution. As a starting point for this work in these experiments, a local measure of self-information of the observed set of intensities at a given location, is used to drive the local deformation. This [12] is given by:

$$\mathcal{S}(\mathbf{x}_R) = -\log p(i'_1, i'_2, \ldots i'_N). \tag{8}$$

This provides a measure of information or uncertainty in terms of the uniqueness of the local set of intensities, which we seek to minimize.

2.4 Geometric Constraints

The above local force alone simply drives intensities together based on their co-occurrence, and the final space they end up in when they coincide is undefined, since there are an infinite number of sets of transformations which bring the structures into correspondence. We therefore require additional constraints to define what transformation solutions are permitted. An ideal solution is to force the estimate to evolve toward the average shape of the brains in the group. Mathematically this implies that, from the reference space $\mathbf{x}_R$, the displacement of each location to each subject, $\mathbf{u}_n(\mathbf{x}_R)$, such that,

$$T_{Rn}(\mathbf{x}_R) = \mathbf{x}_R + \mathbf{u}_n(\mathbf{x}_R) \tag{9}$$

induced by the transformation $T_{Rn}(\mathbf{x}_R)$, summed over the entire group:

$$\mathcal{U}(\mathbf{x}_R) = \left| \sum_{n \in 1 \ldots N} \mathbf{u}_n(\mathbf{x}_R) \right|^2, \tag{10}$$

has zero length, so that each point in the reference space lies at the center of the set of corresponding points in each subject. This term $\mathcal{U}(\mathbf{x}_R)$ can then be used as an additional penalty term in the refinement process.

The second constraint arises from the need to stabilize and regularize the deformation estimate to allow a solution to the 'aperture problem': the requirement of adequate image structure at a point in the scene to support the displacement estimate at that point. There are many approaches to doing this for different registration problems using geometric models from many different sources, including physical material properties, imaging distortion models, or others [10,1]. In this work, to illustrate the general approach of population registration, a simple elastic type deformation energy is used to enforce smoothness of the solution, by quantifying the departure of a single voxel displacement away from the local weighted average of the displacement field around that point. This is given by,

$$\mu_n(\mathbf{x}_R) = \sum_{\mathbf{k} \in K} G(\mathbf{k}) \mathbf{u}_n(\mathbf{x}_R - \mathbf{k}). \tag{11}$$

where $G(\mathbf{k})$ is a low pass spatial kernel (in this case Gaussian) derived from the neighborhood $\mathbf{k} \in K$ around each point $\mathbf{x}$. This is used in the term,

$$\mathcal{R}_n(\mathbf{x}_R) = |\mathbf{u}_n(\mathbf{x}_R) - \mu_n(\mathbf{x}_R)|^2. \tag{12}$$

as a penalty for departure of the point displacement $\mathbf{u}_n(\mathbf{x}_R)$ away from the local regional displacement $\mu_n(\mathbf{x}_R)$. These components, combined with the negative uncertainty in intensity values, provide a force to be minimized to achieve an overall reduction in uncertainty in the co-occurring intensity values:

$$\mathcal{W} = \int_{\mathbf{x}_R} \mathcal{S}(\mathbf{x}_R) + \lambda_1 \int_{\mathbf{x}_R} \mathcal{U}(\mathbf{x}_R) + \lambda_2 \int_{\mathbf{x}_R} \mathcal{R}_n(\mathbf{x}_R), \tag{13}$$

where λ_1 and λ_2 are constants determining the influence of the geometric constraints.

2.5 Implementation Details

The algorithm has been implemented in the form of a dense field volume registration, with a voxel level lattice of displacement estimates $\mathbf{u}_n(\mathbf{x}_R)$. For the experiments presented in the paper, the minimization of (Equation 13) is carried out using a simple iterative scheme consisting of the following steps:

1. Construct a histogram approximation to the joint probability distribution $p(i_1, i_2, \ldots i_n)$ from the set of transformations $\mathcal{T}(\mathbf{x_R})$ by looping over each reference voxel coordinate, using tri-linear interpolation, to estimate the corresponding intensity in each subject image, and updating a count of these held in a sparse N dimensional matrix.
2. Create of a multi-normal parametric approximation to this distribution by placing a Gaussian at each histogram bin with greater than 1 count.
3. Evaluate the intensity gradients $\frac{\partial i'_n(\mathbf{x})}{\partial \mathbf{x}}$ for each subject image for each point in the reference coordinates.
4. Evaluate the change in self information $\mathcal{S}(\mathbf{x}_R)$ induced by a change in the local displacement $\mathbf{u}'_n(\mathbf{x}_R)$.
5. Evaluate $\frac{\partial \mathcal{W}(\mathbf{x}_\mathcal{R})}{\partial \mathbf{u}_n(\mathbf{x}_R)}$ at each location.
6. Update current estimate of $\mathbf{u}'_n(\mathbf{x}_R)$ to minimize $\mathcal{W}$ and goto (1) until convergence.

For these experiments, convergence and termination of the iterations was defined by the fractional reduction in the average value of $\mathcal{W}$ falling below 0.001%.

3 Experimental Investigation

3.1 Synthetic Data

To configure and examine the behavior of the approach to a range of shape variation and object contrasts, an image synthesizer was constructed creating images of spheres of varying contrast, with partial volume model and independent noise. This used a B-Spline based volume deformation [9] algorithm driven

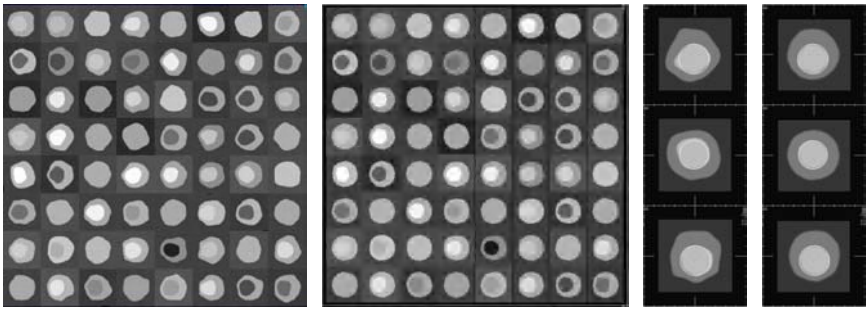


Fig. 3. Left: single axial slices through 64 individuals with random deformation and object contrast relative to base shape. Middle: slices after 35 iterations. Right: orthogonal slices through one individual: deformed individual with overlay of reference shape (left column), and after group spatial normalisation (right column).

by a Gaussian random noise model to create a known shape variation for the spheres. Using this, a population of volumetric images of 64 'subjects', each with 72x72x72 voxels of size 1x1x1 mm was created which are shown in Figure 3.

The results of applying the registration are shown on the right of Figure 3. These verify that the algorithm converges to the mean shape of the population even in the presence of significant misalignment. The registration is also seen to be robust to varying levels of image contrast, recovering the shape and location of the internal sphere, while retaining a meaningful mapping for the sub-group of the population where there is no contrast between the inner and outer sphere.

To evaluate the capability of the algorithm to recover alignment of brain image structure with a wide range of anatomical variation present in normal aging, a set scans of 32 cognitively normal subjects with a range of ages (55 to 77) were selected. Figure 4 illustrates the data and the result in terms of the group averaged image as the iterative alignment proceeds. The Sagittal slice in particular reveals the recovery of sulcal alignment in the cortex.

4 Discussion

This paper has examined some of the key features of a template free approach to population alignment derived from measures of the joint distribution of intensities in the image set being aligned. Differences in the response of this high-dimensional distribution to random population mis-alignment compared to the case of conventional image pair mis-alignment in multi-modality registration were highlighted. These motivated the exploration of a direct approach to population alignment driven by the minimization of the local joint entropy of the combined image values of the population. In addition, by setting up an iterative optimisation scheme that incorporates constraints to force the average displacement field to zero, the algorithm evolves a common mapping relative to the average space of the group of anatomies. Simultaneous multiple image alignment has been examined previously for other applications. Examples from the machine vision community include the recent work of Miller [11] who proposed

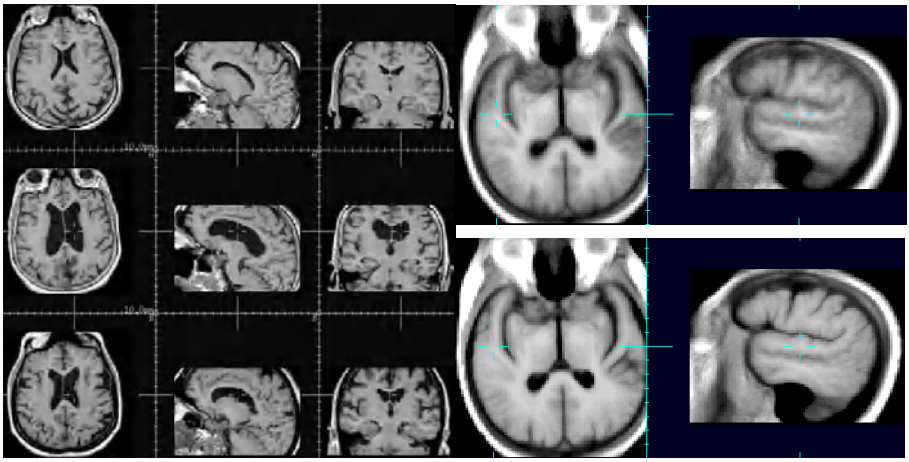


Fig. 4. Left: Example slices through three of the 32 subjects used for experimentation illustrating the range of anatomical variation. Right: Example orthogonal slices through intensity averaged MRI from the group after 1 iteration (bottom) and 40 iterations (top).

an approach to simultaneous binary image alignment using the entropy of the stack of image values at each location. By assuming the images are simple binary values, the method derives an entropy measure separately, from the distribution of the set of values at each location alone, rather than over the entire image. There is therefore no global statistical association derived between intensities in different images.

The approach proposed here contains few priors about the content of the images, regional contrast or modality, which makes it ideal for the study of neuro degenerative processes which induce regional changes in tissue integrity and contrast. Our experiments indicate that the approach seeks a spatial correspondence between common regions visible in sub-groups of the population, while retaining meaningful estimates for those subjects where regions are missing. Computational requirements are, as expected, much higher than simple image pair registration of a population to a template. However, the current implementation in C++ running on a 2.8GHz Pentium machine required only around 1.4GBytes of memory to hold all working data values, and around 4 hours of computing time to align the 64 synthetic subject data. Significant areas for further work are apparent based on this initial experimentation. These include the exploration of the use of other geometric transformation models (such as the fluid transformation [2]), and global entropy criteria such as mutual information [18,3] and normalised mutual information [14], together with evaluation on typical brain image data for different clinical projects.

References

1. G.E. Christensen and H.J. Johnson. Consistent image registration. *Transactions on Medical Imaging*, pages 568–582, 2001.

2. G.E. Christensen, R.D. Rabbitt, and M.I. Miller. Deformable templates using large deformation kinematics. *IEEE Transactions on Image Processing*, 5:1435–1447, 1996.
3. A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multimodality image registration using information theory. In Y. Bizais, C. Barillot, and R. Di Paola, editors, *Proceedings of Information Processing in Medical Imaging*, pages 263–274. Kluwer Academic Publishers, 1995. Brest, France.
4. D.L. Collins, A.C. Evans, C. Holmes, and T.M. Peters. Automatic 3D segmentation of neuro-anatomical structures from MRI. In Y. Bizais, C. Barillot, and R. Di Paola, editors, *Proceedings of Information Processing in Medical Imaging*, pages 139–152. Kluwer Academic Publishers, 1995. Brest, France.
5. C. Davatzikos. Spatial transformation and registration of brain images using elastically deformable models. *Comp Vision and Image Understanding*, 66(2):207–222, 1997.
6. J.C. Gee, M. Reivich, and R. Bajcsy. Elastically deforming 3d atlas to match anatomical brain images. *J. Comput. Asst Tomogr*, 17:225–236, 1993.
7. U. Grenander and M. I. Miller. Computational anatomy: An emerging discipline. *Quart. Appl. Math*, 56(4):617–694, 1998.
8. D.L.G. Hill, C. Studholme, and D.J. Hawkes. Voxel similarity measures for automated image registration. In *Proceedings of Visualisation in Biomedical Computing*, volume SPIE 2359, pages 205–216. SPIE Press, 1994. Rochester Mn.,U.S.A.
9. S. Lee, G. Wolberg, and S.Y. Shin. Scattered data interpolation with multilevel B-splines. *IEEE Transactions on Visualisation and Computer Graphics*, 3:228–244, 1997.
10. H. Lester, S. Arridge, K. Jansons, L. Lemieux, J. Hajnal, and A. Oatridge. Non-linear registration with variable viscosity fluid algorithm. In *Proc. Image Processing in Medical Imaging*, pages 238–251. Springer Verlag, 1999.
11. E. Miller. *Learning from One Example in Machine Vision by Sharing Probability Densities*. PhD thesis, Massachusetts Institute of Technology, 2002.
12. F.M. Reza. *An Introduction to Information Theory*. Dover, New York, 1994.
13. C. Studholme, V. Cardenas, N. Schuff, H. Rosen, B. Miller, and M. Weiner. Detecting spatially consistent structural differences in alzheimer's and fronto temporal dementia using deformation morphometry. In *Proceedings of 4th International Conference on Medical Image Computing and Computer Assisted Interventions, 2001*, pages 41–48, 2001.
14. C. Studholme, D.L.G. Hill, and D.J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1):71–86, 1999.
15. G. Subsol, J.-P. Thirion, and N. Ayache. A scheme for automatically building three-dimensional morphometric anatomical atlases: Application to a skull atlas. *Medical Image Anal.*, 2:37–60, 1998.
16. J.-P. Thirion. Image matching as a diffusion process: An analogy with maxwell's demons. *Medical Image Anal.*, 2(3):243–260, 1998.
17. P. Thompson, R. Woods, M. Mega, and A. Toga. Mathematical/Computational challenges in creating deformable and probabilistic atlases of the human brain. *Human Brain Mapping*, 9:81–92, 2000.
18. P.A. Viola and W.M. Wells. Alignment by maximisation of mutual information. In *Proceedings of the 5th International Conference on Computer Vision*, pages 15–23, 1995.
19. R.P. Woods, S.T. Grafton, J.D. Watson, N.L. Sicotte, and J.C. Mazziotta. Automated image registration II: Intersubject validation of linear and non-linear models. *Journal of Computer Assisted Tomography*, 22:153– 165, 1998.

Piecewise Affine Registration of Biological Images

Alain Pitiot[1,2], Grégoire Malandain[1],
Eric Bardinet[1,3], and Paul M. Thompson[2]

[1] Epidaure, INRIA, 2004 route des lucioles BP 93, 06 902 Sophia-Antipolis, France
[2] LONI, UCLA School of Medicine, Los Angeles, CA 90095, USA
[3] CNRS UPR640-LENA, 47 blvd. de l'hôpital, 75651 Paris, France

Abstract. This manuscript tackles the registration of 2D biological images (histological sections or autoradiographs) to 2D images from the same or different modalities (*e.g.*, histology or MRI). The process of acquiring these images typically induces composite transformations that can be modeled as a number of rigid or affine local transformations embedded in an elastic one. We propose a registration approach closely derived from this model. Given a pair of input images, we first compute a dense similarity field between them with a block matching algorithm. A hierarchical clustering algorithm then automatically partitions this field into a number of classes from which we extract independent pairs of sub-images. Finally, the pairs of sub-images are, independently, affinely registered and a hybrid affine/non-linear interpolation scheme is used to compose the output registered image. We investigate the behavior of our approach under a variety of conditions, and discuss examples using real biomedical images, including MRI, histology and cryosection data.

1 Introduction

A key component of medical image analysis, image registration essentially consists of bringing two images, acquired from the same or different modalities, into spatial alignment. For instance, monomodal registration of a population's MRIs can be used to build anatomical atlases [1,2], while mono- or multi-modal registration of the same patient's data can help determine the nature of an anomaly or monitor the evolution of a tumor or other disease process.

In particular, pair-by-pair registration of a series of 2-D biological images (histological sections or autoradiographs) enables the reconstruction of a 3D biological image. Subsequent fusion with 3D data acquired from tomographic imaging modalities (*e.g.* MRI) then allows the tissue properties to be studied in an adequate anatomic framework, using *in vivo* reference data.

More formally, given two input images, registering the floating (*i.e.*, moveable) image to the reference (*i.e.*, fixed) one entails finding the transformation that minimizes the dissimilarity between the transformed floating image and the reference. As such, it can be decomposed into 3 elements: a transformation space, a similarity metric and an optimization algorithm (see [3] for an extensive review

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 91–101, 2003.

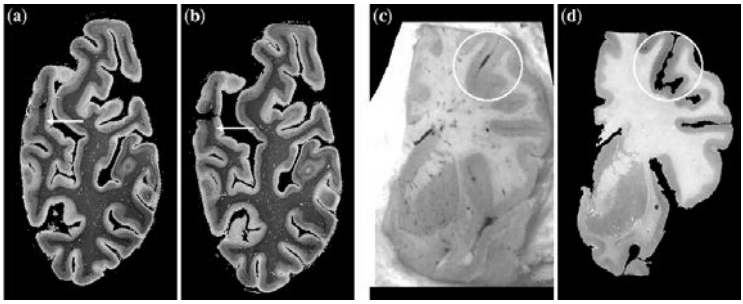


Fig. 1. Two consecutive myelin-stained histological sections of the human brain (a & b); Human brain cryosection (c) and its associated Nissl-stained section (d).

of medical image registration). For iconic (*i.e.*, intensity-based) methods, optimal similarity measures can be derived from a careful analysis of the expected relationships between the input images [4], with different hypotheses leading to different measures. Similarly, *a priori* knowledge about the acquisition process for biological images may allow the transformation space to be modeled more accurately.

In our case, the cutting process, successive chemical treatments, and the glass mounting step that a slab of tissue undergoes during a histological preparation yield a fairly flexible global transformation that is however locally affine for some identifiable components of the section. In brain sections for instance, each gyrus (compare white arrows in Figure 1.a & b, and white circles in c & d) undergoes an affine transformation (due to successive manipulations) relatively independent from those of other gyri. Consequently, even though a large variety of transformation spaces have been discussed in the literature (among others, one finds linear (rigid, affine) and non-linear transformations (polynomial [5], elastic [6,7] or fluid [8]), their functional form may not reflect our specific needs.

Note that the utility of the above-stated transformation model extends to medical as well as biological images (our primary motivation here). For instance, abdominal or torso MRIs often include rigid structures such as bones (ribs, vertebrae, etc.), deformable organs (liver, heart, etc.), and elastic tissues. Two abdominal MRIs of the same patient are then linked by a complex transformation which can be rigid in some regions (for bones) but potentially exhibits large local dilations (in deformable organs). Global rigid or affine transformations obviously cannot handle such a case adequately. Also, a single rigid transformation would not correctly register all the vertebrae along the spinal column simultaneously. Furthermore, high degree of freedom (*e.g.*, fluid) transformations could correctly map one image onto the other, but they may not ensure that specific components (*e.g.*, bones) will be only rigidly transformed.

To alleviate these issues, a few authors have developed local registration techniques, where the input images are divided into a number of smaller sub-images, and a transformation is associated with each. An automatic hierarchical elastic image registration technique is presented in [9]. The initial 2-D images are partitioned into quadtree structures. At each level of the quadtree, the floating

sub-images are independently registered to their counterparts in the reference image, before being merged via thin-plate spline interpolation. However, this technique cannot apply a transform selectively to a specific region whose boundary does not coincide with the quadtree grid. In [10], Little *et al.* describe an approach where a user selects a number of pairs of corresponding rigid structures in the input images along with associated linear transformations (also given by the user). A number of pairs of landmarks further constrain a hybrid affine/non-linear interpolation scheme that acts as a local registration algorithm. This method, even though it fits our needs, still relies on interactive specification of the components to be rigidly matched.

This paper addresses the problem of *automatically* registering two images, when the images consist of a number of independent components, subject to linear transformations. Figure 2 illustrates our approach. First a correspondence field (section 2.1) is computed between the two images. From a hierarchical clustering of this field (section 2.2), we automatically extract pairs of sub-images (with an assumed linear transformation relation between the pairs). These pairs are then, separately, rigidly or affinely registered. Finally, the hybrid affine/non-linear interpolation scheme described in [10] is used to compose the registered floating image (section 2.3). Results on phantom and real data are presented in section 3.

2 Method

The first step of our approach consists of *automatically* partitioning the input floating and reference images, I_F and I_R, into a number of pairs of corresponding sub-images, where each sub-image is associated with an independent image component (in terms of transformation). We approach this segmentation issue as a process of partitioning a correspondence field computed from I_F to I_R. Our method is motivated by the following observation. When both images are composed of pairs of independent components, where each component is subject to some linear transformation, the associated correspondence field should exhibit rather homogeneous characteristics within each component, and heterogeneous ones across them.

2.1 Computing the Correspondence Field

We use a block-matching algorithm [11] to compute the correspondence field. We associate with I_F and I_R two sparse regular rectangular lattices $L_F = [1 \dots w_F] \times [1 \dots h_F]$ and $L_R = [1 \dots w_R] \times [1 \dots h_R]$. From these we discard, for histological sections, sites that lie on the background. Note that we may choose to associate a site to each pixel of the input images, in which case w_F, h_F and w_R, h_R would be the width and height of I_F and I_R.

The correspondence field then associates to each site (i, j) in L_F a 2-D spatial similarity distribution and an "optimal" displacement, computed as follows. For each site (i, j) in L_F, we consider in I_F a neighborhood $B^{i,j}_{I_F}$ of the pixels associated with (i, j) (usually a square neighborhood of constant size called

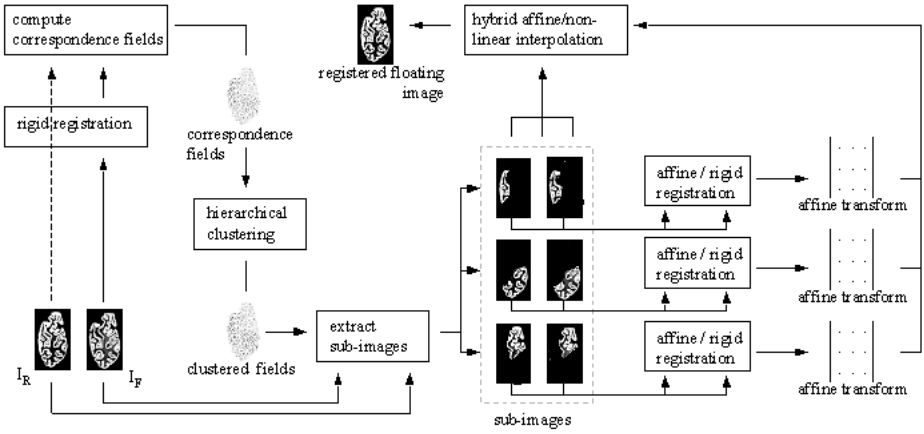


Fig. 2. Overview of our local registration approach.

a "block"), whose centroid is denoted by $p_F^{i,j}$. We then compute the similarity measures (given a similarity metric *sim*) between block $B_F^{i,j}$ and every block $B_R^{k,l}$ in I_R associated to sites (k,l) in the corresponding neighborhood $N_{i,j}$ of (i,j) in L_R (the "exploration neighborhood"). For every site (i,j) in L_F, we then get a 2-D spatial similarity distribution (the values $sim(B_F^{i,j}, B_R^{k,l})$ defined in the neighborhood $N_{i,j}$ of (i,j) in L_R), and an "optimal" displacement $d^{i,j}$ defined by $d^{i,j} = p_R^{(k,l)_{max}} - p_F^{i,j}$ where $(k,l)_{max}$ is the site of L_R that is associated to the block $B_R^{k,l}$ in $N_{i,j}$ which is the most similar to $B_F^{i,j}$, *i.e.* $(k,l)_{max} = \arg\max_{k,l} sim(B_F^{i,j}, B_R^{k,l})$.

The quality of both the similarity distribution and the displacement field is essentially determined by three parameters: the block size, the similarity metric and the size of the exploration neighborhood in L_R. The similarity metric and block size must reflect the expected relationship between the intensity distributions of blocks in the floating and reference images, and the scale of the features of interest within those blocks [12,1]. For most of our experiments, we used the correlation coefficient [13] (which assumes an affine relation between the intensity distributions of the blocks). The size of the exploration neighborhood is linked to the expected magnitude of the residual displacements after global alignment. It conditions the extent to which our registration algorithm can recover large deformations. In our case, it is input by the user (typically 15 pixels here).

As a pre-processing step, we first rigidly register I_F to I_R to remove the global rigid transform (that uniformly affects all components) from the subsequently computed correspondence fields. We use the fully automated intensity-based registration algorithm presented in [11], where a *robust* multi-scale block-matching strategy was introduced.

2.2 Extracting the Image Components

Clustering the Correspondence Field. We are looking for a *hierarchical clustering* of L_F, that is, a sequence of partitions in which each partition is nested into the next partition in the sequence [14]. Cluster analysis (unsupervised learning) essentially consists of sorting a series of multi-dimensional points into a number of groups (clusters) to maximize the intra-cluster degree of association and minimize the inter-cluster one. It is particularly well-suited here as it behaves adequately even when very little is known about the category structure of the input set of points.

Our clustering method is adapted from the standard agglomerative hierarchical clustering algorithm [15]:

step 1: initialize a cluster list by placing each site of L_F in an individual cluster, and let the distance between any two of these clusters be the distance between the sites they contain.
step 2: find the closest pair of clusters, remove them from the cluster list, merge them into a new single cluster and add the new cluster to the cluster list.
step 3: compute the distances between the newly formed cluster and the other ones in the cluster list.
step 4: repeat steps 2 and 3 until the desired number of clusters is reached.

The number of clusters can either be specified by the user (our case here), or by using pre-indicators like the Davies-Bouldin index [16] or the cophenetic correlation coefficient [14] to assist this choice.

Computation of these cluster distances is the pivotal element of the clustering algorithm. The distance should be consistent with the model we choose for the input images and the relationships we expect between them.

To define a distance on clusters, we first need a distance on sites. Given two sites, t and u, their distance is defined as a linear combination of three distances:

$$D_{site}(t,u) = \alpha\ D_p(p_F^t, p_F^u) + \beta\ D_d(d^t, d^u) + \gamma\ D_\rho(\rho^t, \rho^u) \tag{1}$$

$D_p(p_F^t, p_F^u)$ is the geodesic distance between the block centroids. It ensures that close blocks are more likely to be clustered than blocks that are far apart (model constraint), while adequately representing proximity from an anatomical point of view when the input images contain several pieces of tissue.

$D_d(d^t, d^u)$ is the distance between the block displacement vectors, we choose the modulus of their difference, *i.e.* $\|d^t - d^u\|$.

$D_\rho(\rho^t, \rho^u)$ is the distance between the block similarity distributions. Clearly, the "optimal" displacement may sometimes disagree with the actual movement (because of noise, decoys, etc.). The similarity distribution is then better-suited to capture fine structural elements and discriminate between blocks. We use a normalized version ρ of these distributions to ensure that they all have the same overall unit mass. For each site t in L_F and each site u in its exploration neighborhood, ρ is given by: $\rho^t(p_R^u - p_F^t) = sim(B_F^t, B_R^u) / \sum_v sim(B_F^t, B_R^v)$. As a distance between distributions, we chose the Earth mover's distance [17], a

discrete solution to the discrete Monge-Kantorovich mass-transfer problem [18]. Given the "ground distance" (the distance between elements of the distribution, the Euclidean distance in our case), the Earth mover's distance (EMD) between two distributions becomes the minimal total amount of work (= mass × distance) it takes to transform one distribution into the other. As argued by Rubner *et al.* [17], this boils down to a bipartite network flow problem, which can be modeled with linear programming and solved by a simplex algorithm.

α, β and γ are real-valued positive weights. We choose for β a value substantially smaller than that of α and γ (typically, $\beta = \frac{\alpha}{50}$).

Once we have a distance between blocks, a cluster distance can be defined. We adapted the standard complete link distance [14] to additionally take into account the transformations that can be estimated on the clusters already formed. Namely, when the size of a cluster reaches a given threshold (we usually take 20, even though experiments showed that the value of that threshold does not really impact the quality of the clustering), a rigid or affine transformation can be estimated, in a robust fashion, from the associated set of "optimal" displacement vectors (*e.g.*, the quaternion-based approach presented in [19]). The decision to merge two clusters can then be biased by the agreements between the directions of the "optimal" displacement vectors of one cluster with the estimated transformation of the other as this might indicate that they belong to the same component.

Given a site t in L_F, its distance to a transformation T is defined by the coherence between T and the "optimal" displacement vector associated with t. Consequently, the transformation-based distance between a site t and a cluster C is defined by the distance between t and T^C, the transformation estimated over C: $D_{tran}(t, C) = \|T^C(p_F^t) - (p_F^t + d_F^t)\|$ if T^C is defined, and 0 otherwise.

Given two clusters of sites $C^a = \{a_1, \ldots, a_{n_a}\}$ and $C^b = \{b_1, \ldots, b_{n_b}\}$, with associated estimated transformations T^a and T^b respectively, the cluster distance between them is defined as the longest distance from any block of C^a to any block of C^b (complete-link) *plus* the sum of the "transformation distances" wherever they can be computed:

$$D_{cluster}(C^a, C^b) = \max_{i,j} D_{site}(a_i, b_j) + \sum_j D_{tran}(b_j, C^a) + \sum_i D_{tran}(a_i, C^b) \quad (2)$$

Extracting the Sub-images. Let N_C be the final number of clusters, $C = \{C^1, \ldots, C^{N_C}\}$ the cluster partition of L_F, and $\{c_1^i, \ldots, c_{n_i}^i\}$ the n_i sites of the i^{th} cluster C^i. We want to build a set of N_C sub-images $\{I_F^i\}_{i=1}^{N_C}$, each of them associated with a single cluster. Our clustering method does not ensure that the borders between clusters are sufficiently precise to adequately represent the sub-images' borders. Moreover, as we are going to use these sub-images to find local transformations, it is often better to choose larger supports to avoid boundary effects.

Consequently, rather than build a partition of I_F from the partition of L_F, we build a covering of I_F, *i.e.*, a set of sub-images that could overlap. To do so, we aggregate in I_F^i the pixels of I_F in the vicinity of the sites of the cluster C^i.

We get: $I_F^i = \{(x,y) \in I_F \text{ such that } D((x,y), p_F^{c_j^i}) \leq radius \text{ for some } c_j^i \in C^i\}$ In practice we use the L_∞ distance. Then, if the blocks associated to the sites are of size $n \times n$, taking $radius = n/2$ we get $I_F^i = \bigcup_j B_{I_F}^{s_j^i}$. In our experiments, to ensure a large support for the sub-images, we chose $radius = 3n/4$.

The corresponding reference sub-images I_R^i are built identically, but with the centroids $p_R^{(k,l)_{max}}$ of the most similar blocks (see Section 2.1): $I_R^i = \{(x,y) \in I_R \text{ such that } D((x,y), p_F^{c_j^i} + d^{c_j^i}) \leq radius, \text{ for some } c_j^i \in C^i\}$. Again, we use the L_∞ distance here, with $radius = n$ (a larger extent than that of the floating sub-image) to ensure that I_F^i can be effectively registered against I_R^i.

2.3 Composing the Registered Floating Image

Once we have extracted the reference and floating sub-images, we use the robust block-matching algorithm described in [11] to register them, independently, pair by pair. For each pair $\left\{I_R^l, I_F^l\right\}, l \in 1 \ldots N_C$, we obtain an affine transform T^l. Note that as these registrations are robust, the sub-images do not need to perfectly correspond to the anatomically separate components.

We then compose the final registered floating image using the Little *et al.* method (see [10] for details). Their approach applies user-provided affine transforms to user-defined structures and ensures a smooth interpolation in between them. In our application, the set of floating sub-images forms a covering of the input floating image, so we have to erode the sub-images to leave space for interpolation. Furthermore, the floating sub-images must be cut to ensure that they do not overlap, once transformed, as this may also impair the interpolation scheme.

Note that the entire registration process could easily be included within an iterative multi-scale framework to achieve a better trade-off between accuracy and complexity. Such a framework could also handle both large-scale and small-scale components.

3 Results

3.1 Synthetic Experiments

Synthetic fields enable us to evaluate the quality of the clustering algorithm independently from textural issues and aside of considerations regarding the similarity distribution. That is, in simulated data, we can assume that the "optimal" displacement field is the correct one. Due to a lack of space, we outline the quantitative results only briefly here. Please refer to
`http://www-sop.inria.fr/epidaure/personnel/apitiot/WBIR/index.html`
for the corresponding figures and tables.

The first experiment was designed to illustrate how our approach behaves on images consisting of several *separate* independent components. We considered a spine-like structure with 4 components (4 squares) to which 4 different affine transformations were applied. We computed for each component an exact field and then corrupted it with a uniform noise: for each component, $\alpha\%$ of

the displacement vectors were selected for perturbation. Our algorithm managed to perfectly cluster the components even for very high values of noise ($\alpha = 75\%$), mostly due to the use of the geodesic distance. Additionally, the estimated transformations were within 5% of the desired ones up until $\alpha = 75\%$. A second experiment demonstrated the behavior of our approach for a *connected* component. We used similar settings for 3 more structures with 2 components each (a small square inside a larger one, and two vertical halves of a large square). Since the geodesic distance did not help here, we obtained inferior performances (20% more wrongly clustered vectors at $\alpha = 50\%$). This may also be because the displacement vectors were quite similar on both sides of the border between the components. Throughout, clustering results with the transformation-based distance were consistently better than without, and even more so when the noise level increased (40% more wrongly clustered vectors at $\alpha = 25\%$, 50% at $\alpha = 75\%$). Incidentally, a perfect clustering is not necessary since the subsequently extracted sub-images are "larger" than the clusters and their independent registrations are robust.

We used phantom images and the same experimental settings to evaluate how our clustering algorithm performed under controlled textural conditions. Namely, pairs of synthetic images were created with artifical textures (grayscale cloud patterns) and the same transformations applied as above. We observed better performances when the EMD distribution distance was used than when it was not (15% better on average). The similarity distribution distance helps the clustering algorithm to form, at early stages, sensible clusters. These are then agglomerated with the aid of the robust estimation of the associated transformations. Again, results were better with the cluster transformations estimation (9% better). However the difference in performance was less obvious than before: as the optimal displacement field is very noisy, the transform distance only marginally helps the clustering process. A more thorough validation using complete series of histological sections is in progress.

3.2 Biomedical Images

Figure 3 displays the results of our local registration for a pair of myelin-stained histological sections (myelin-stained coronal section through the occipital cortex, first row), and for a Nissl stained human brain section and its associated cryosection image (second row). It demonstrates the behavior of the piecewise approach (1) when reconstructing a histological block from a series of stained sections, and (2) in a multi-modal registration problem. These examples present two classic registration difficulties: in the first case, a gyrus (top left corner) was detached during the histological preparation and manually realigned in an unsatisfactory fashion; in the second case, many gyri were separated during histological treatment.

The reference images are those of Figure 1. For each pair of images, we show in Figure 3 the reference image with the superimposed edges of the globally affinely registered floating image (a), the floating image with clustered optimal displacement field (b), the image of a regular grid convected by the associated

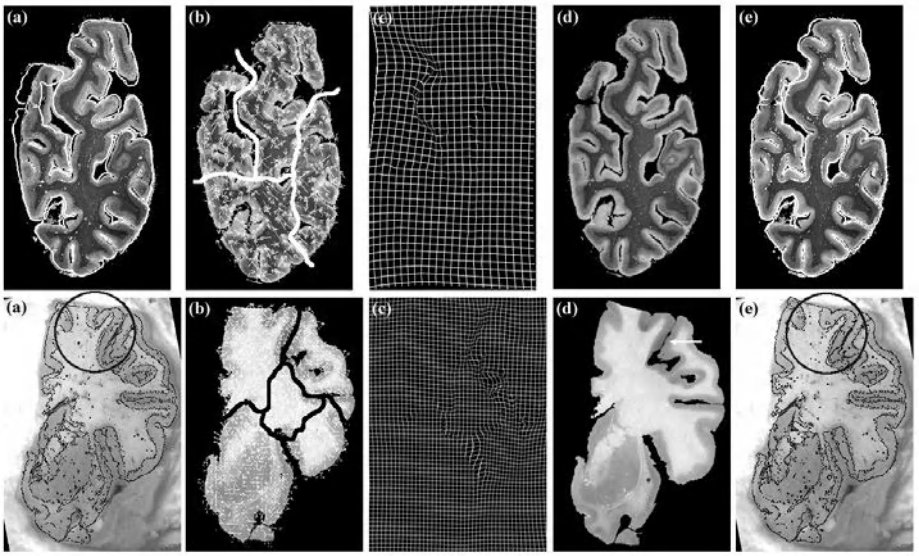


Fig. 3. Registration of two consecutive myelin-stained histological sections of the human brain (first row) and of a Nissl-stained human brain section and its associated cryosection (second row).

hybrid affine/non-linear transformation (c), the locally registered floating image (d), and the reference image with the superimposed edges of the locally registered floating image (e).

To better demonstrate how our approach can register multiple component images, we applied only rigid transforms to the sub-images of the myelin-stained sections. The clustering algorithm adequately isolated in a separate sub-image the floating gyrus which was subsequently correctly registered to its counterpart in the reference image. Our technique also successfully compensated for smaller rotations applied by the operator in the glass mounting step. An affine transform would of course further decrease the discrepancy between the pairs of sub-images. However, in the general case, when one suspects only a rigid transformation between sub-images, opting for an affine registration would only introduce unnecessary over-parameterization. Among other disadvantages this could substantially alter textures.

For the multi-modal example, we used affine registration and specified a larger number of clusters (6) to take into account the many inadequately manually realigned components. Our algorithm successfully closed most of the gyri (see white arrow in Figure 3.d and compare circled area in 3.a and 3.e) while minimizing the overall amount of deformation.

4 Conclusion and Perspectives

We have presented a fully automated local registration method, capable of dealing with a variety of 2-D images. It builds complex spatial transformations by

elastically interpolating between rigid or affine transforms that are locally defined on pairs of sub-images. Consequently, we minimize the overall number of degrees of freedom of the transformation, following the guidelines of the parsimony principle (see [20] for a discussion of some issues associated with high-dimensional transformations). These sub-images represent geometrically coherent components (in our biomedical applications, they are even anatomically coherent components). They are automatically extracted from an initial correspondence field computed between the images. All user interaction is avoided, by contrast with other approaches [10].

The use of a hierarchical clustering approach and a similarity distribution distance proved promising: while the distribution distance can effectively deal with noise and textural issues to discriminate between image blocks, our clustering algorithm manages to extract the expected sub-images.

Results on real data showed that the proposed method is adequate for several specific problems in biomedical imaging. Finally, even though the presented method works in 2-D, it could readily be extended to 3-D (or n-D) with a close-to-linear increase in processing time.

References

1. Collins, D.L., Zijdenbos, A.P., Paus, T., Evans, A.C.: Use of registration for cohort studies. In Hajnal, J., Hawkes, D., Hill, D., eds.: Medical Image Registration. (2003)
2. Thompson, P.M., Woods, R.P., Mega, M.S., Toga, A.W.: Mathematical/ computational challenges in creating deformable and probabilistic atlases of the human brain. Human Brain Mapping **9** (2000) 81–92
3. Maintz, J.B.A., Viergever, M.A.: A survey of medical image registration. Medical Image Analysis **2** (1998) 1–36
4. Roche, A., Malandain, G., Ayache, N.: Unifying Maximum Likelihood Approaches in Medical Image Registration. International Journal of Imaging Systems and Technology: Special Issue on 3D Imaging **11** (2000) 71–80
5. Woods, R.P., Grafton, S.T., Holmes, C.J., Cherry, S.R., Mazziotta, J.C.: Automated image registration: I. General methods and intrasubject, intramodality validation. J. Comput. Assist. Tomogr. **22** (1998) 141–154
6. Davatzikos, C.: Spatial transformation and registration of brain images using elastically deformable models. CVIU **66** (1997) 207–222
7. Gee, J.C., Reivich, M., Bajcsy, R.: Elastically Deforming 3D Atlas to Match Anatomical Brain Images. J. Comput. Assist. Tomogr. **17** (1993) 225–236
8. Christensen, G.E.: Consistent linear-elastic transformations for image matching. In: Proc. of IPMI. Volume 1613 of LNCS., Springer (1999) 224–237
9. Likar, B., Pernus, F.: Registration of Serial Transverse Sections of Muscular Fibers. Cytometry **37** (1999) 93–106
10. Little, J.A., Hill, D.L.G., Hawkes, D.J.: Deformations Incorporating Rigid Structures. Computer Vision and Image Understanding **66** (1997) 223–232
11. Ourselin, S., Roche, A., Subsol, G., Pennec, X., Ayache, N.: Reconstructing a 3D Structure from Serial Histological Sections. Image and Vision Computing **19** (2001) 25–31

12. Dengler, J.: Estimation of discontinuous displacement vector fields with the minimum description length criterion. In: Proc. of CVPR, IEEE (1991) 276–282
13. Penney, G.P., Weese, J.W., Little, J.A., Desmedt, P., Hill, D.L.G., Hawkes, D.J.: A comparison of similarity measures for use in 2D-3D medical image registration. In: Proc. of MICCAI. Volume 1496 of LNCS., Springer (1998) 1153–1161
14. Backer, E.: Computer-assisted reasoning in cluster analysis. Prentice Hall (1995)
15. Johnson, S.C.: Hierarchical Clustering Schemes. Psychometrika **2** (1967) 241–254
16. Davies, D.L., Bouldin, D.W.: A cluster separation measure. IEEE Trans. PAMI **1** (1979) 224–227
17. Rubner, Y., Tomasi, C., Guibas, L.: A metric for distributions with applications to image databases. In: Proc. of ICCV, Bombay, India, IEEE (1998)
18. Haker, S., Angenent, S., Tannenbaum, A.: Minimizing Flows for the Monge-Kantorovich Problem. SIAM Journal of Math Analysis (2003) [to appear].
19. Horn, B.K.P.: Closed-form solution of absolute orientation using unit quaternions. Journal of the Optical Society of America A **4** (1987) 629–642
20. Kjems, U., Hansen, L.K., Chen, C.T.: A non-linear 3d brain co-registration method. In Hansen, P.C., ed.: Proceedings of the Interdisciplinary Inversion Workshop 4, Lyngby, Denmark, IMM, Technical University of Denmark (1996)

Elastic Registration with Partial Data*

Senthil Periaswamy and Hany Farid

Dartmouth College, Hanover, NH, 03755, USA

Abstract. We have developed a general purpose registration algorithm for medical images and volumes. The transformation between images is modelled as locally affine but globally smooth, and explicitly accounts for local and global variations in image intensities. An explicit model of missing data is also incorporated, allowing us to simultaneously segment and register images with partial or missing data. The algorithm is built upon a differential multiscale framework and incorporates the expectation maximization algorithm. We show that this approach is highly effective in registering a range of synthetic and clinical medical images.

1 Introduction

The goal of image registration is to find a transformation that aligns one image to another. Medical image registration has emerged from this broad area of research as a particularly active field (see, for example, [9,7] for general surveys). This activity is due in part to the many clinical applications including diagnosis, longitudinal studies, and surgical planning, and to the need for registration across different imaging modalities (e.g., MRI, CT, PET, X-RAY, etc.). Medical image registration, however, still presents many challenges. Several notable difficulties are 1.) the transformation between images can vary widely and be highly nonlinear (elastic) in nature; 2.) images acquired from different modalities may differ significantly in overall appearance and resolution; 3.) there may not be a one-to-one correspondence between the images (missing/partial data); and 4.) each imaging modality introduces its own unique challenges, making it difficult to develop a single generic registration algorithm.

In our earlier work, we described a general-purpose registration algorithm that contends with both large- and small-scale geometric and intensity distortions [10]. In this paper, we describe an extension to this work that allows us to explicitly contend with missing or partial data. Shown in Fig. 2 are examples of the challenges posed by missing data. In these examples there are large portions of the source image that have no corresponding match in the target image. Without an explicit segmentation or localization of these missing regions, most registration algorithms are unlikely to correctly register these images. Of course, if the registration between these images were known, then it would be straightforward to perform the segmentation. Similarly, if the segmentation were known,

* This work was supported by an Alfred P. Sloan Fellowship, a NSF CAREER Award (IIS-99-83806), and a department NSF infrastructure grant (EIA-98-02068). The authors can be reached at `sp@cs.dartmouth.edu` and `farid@cs.dartmouth.edu`.

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 102–111, 2003.

the registration could proceed. Without a known segmentation or registration, however, we are faced with a bit of a chicken and egg problem - which step should be performed first? In order to contend with this problem we have employed the expectation maximization algorithm that simultaneously segments and registers a pair of images or volumes (see also [2]).

For purposes of completeness we will briefly review our previous registration algorithm [10], and then describe the extension that allows us to contend with missing data. We then show the efficacy of this approach on several synthetic and clinical cases.

2 Registration

We formulate the problem of image registration within a differential (non feature-based) framework. This formulation borrows from various areas of motion estimation (e.g., [6,3]). In order to contend with partial or missing data, the expectation maximization algorithm [4] is incorporated into this framework, allowing for simultaneous segmentation and registration. We first outline the basic computational framework, and then discuss several implementation details that are critical for a successful implementation.

2.1 Local Affine

Denote $f(x, y, t)$ and $f(\hat{x}, \hat{y}, t-1)$ as the source and target images, respectively [1]. We begin by assuming that the image intensities between images are conserved (this assumption will be relaxed later), and that the geometric transformation between images can be modeled locally by an affine transform:

$$f(x, y, t) = f(m_1 x + m_2 y + m_5, m_3 x + m_4 y + m_6, t - 1), \tag{1}$$

where m_1, m_2, m_3, m_4 are the linear affine parameters, and m_5, m_6 are the translation parameters. These parameters are estimated locally for each small spatial neighborhood, but for notational convenience their spatial parameters are dropped. In order to estimate these parameters, we define the following quadratic error function to be minimized:

$$E(\boldsymbol{m}) = \sum_{x,y \in \Omega} \left[f(x, y, t) - f(m_1 x + m_2 y + m_5, m_3 x + m_4 y + m_6, t - 1) \right]^2, \tag{2}$$

where $\boldsymbol{m} = (\, m_1 \quad \ldots \quad m_6 \,)^T$, and Ω denotes a small spatial neighborhood. Since this error function is non-linear in its unknowns, it cannot be minimized analytically. To simplify the minimization, we approximate this error function using a first-order truncated Taylor series expansion:

[1] We adopt the slightly unconventional notation of denoting the source and target image with a temporal parameter t. This is done for consistency within our differential formulation.

$$E(\boldsymbol{m}) \approx \sum_{x,y\in\Omega} [f_t(x,y,t) - (m_1 x + m_2 y + m_5 - x) f_x(x,y,t) \\ -(m_3 x + m_4 y + m_6 - y) f_y(x,y,t)]^2 , \quad (3)$$

where $f_x(\cdot)$, $f_y(\cdot)$, $f_t(\cdot)$ are the spatial/temporal derivatives of $f(\cdot)$. Note that this quadratic error function is now linear in its unknowns, $\boldsymbol{m}$. This error function may be expressed more compactly in vector form as:

$$E(\boldsymbol{m}) = \sum_{x,y\in\Omega} \left[k - \boldsymbol{c}^T \boldsymbol{m}\right]^2 , \quad (4)$$

where the scalar k and vector $\boldsymbol{c}$ are given as: $k = f_t + x f_x + y f_y$ and $\boldsymbol{c} = (\, x f_x \quad y f_x \quad x f_y \quad y f_y \quad f_x \quad f_y \,)^T$. This error function can now be minimized analytically by differentiating with respect to the unknowns:

$$\frac{dE(\boldsymbol{m})}{d\boldsymbol{m}} = \sum_{x,y\in\Omega} -2\boldsymbol{c}\left[k - \boldsymbol{c}^T \boldsymbol{m}\right] , \quad (5)$$

setting the result equal to zero, and solving for $\boldsymbol{m}$ yielding:

$$\boldsymbol{m} = \left[\sum_{x,y\in\Omega} \boldsymbol{c}\, \boldsymbol{c}^T\right]^{-1} \left[\sum_{x,y\in\Omega} \boldsymbol{c}\, k\right] . \quad (6)$$

This solution assumes that the first term, a 6×6 matrix, is invertible. This can usually be guaranteed by integrating over a large enough spatial neighborhood Ω with sufficient image content. When an estimate cannot be made, the local parameters are interpolated from nearby regions. With this approach a dense locally affine mapping can be found between a source and target image.

2.2 Intensity Variations

Inherent to the model outlined in the previous section is the assumption that the image intensities between the source and target are unchanged (brightness constancy). This assumption is likely to fail under a number of circumstances. To account for intensity variations, we incorporate into our model an explicit change of local contrast and brightness [12]. Specifically, our initial model, Equation (1), now takes the form:

$$m_7 f(x,y,t) + m_8 = f(m_1 x + m_2 y + m_5, m_3 x + m_4 y + m_6, t-1), \quad (7)$$

where m_7 and m_8 are two new (also spatially varying) parameters that embody a change in contrast and brightness, respectively. Note that these parameters have been introduced in a linear fashion. As before, this error function is approximated with a first-order truncated Taylor series expansion to yield:

$$E(\boldsymbol{m}) = \sum_{x,y\in\Omega} \left[k - \boldsymbol{c}^T \boldsymbol{m}\right]^2 , \quad (8)$$

where the scalar k and vector $\boldsymbol{c}$ are now given as:

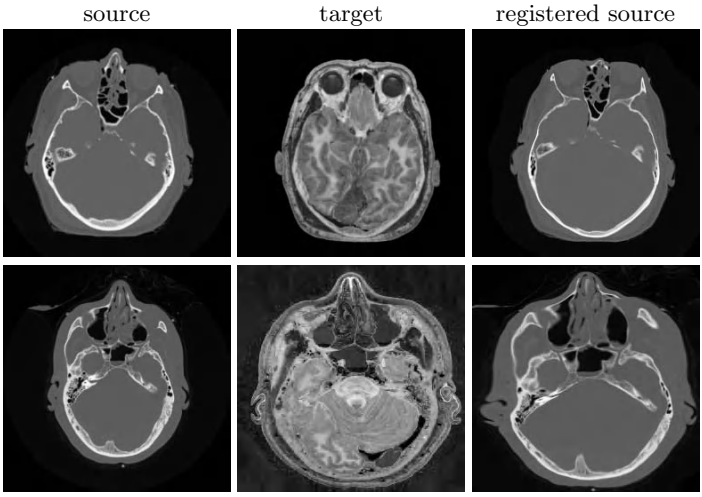


Fig. 1. Shown are examples of registration in the presence of significant geometric and intensity variations. The source are CT images, and the target are photographs from the Visible Human Project.

$$k = f_t - f + xf_x + yf_y \tag{9}$$

$$\boldsymbol{c} = (\, xf_x \quad yf_x \quad xf_y \quad yf_y \quad f_x \quad f_y \quad -f \quad -1\,)^T, \tag{10}$$

Minimizing this error function is accomplished as before by differentiating $E(\boldsymbol{m})$, setting the result equal to zero and solving for $\boldsymbol{m}$. The solution takes the same form as in Equation (6), with k and $\boldsymbol{c}$ as defined in Equations (9) and (10).

Intensity variations are typically a significant source of error in differential motion estimation. The addition of the contrast and brightness terms allows us to accurately register images in the presence of these variations, Fig. 1.

2.3 Smoothness

Until now, we have assumed that the local affine and contrast/brightness parameters are constant within a small spatial neighborhood, Equation (8). There is a natural trade-off in choosing the size of this neighborhood. A larger area makes it more likely that the matrix in Equation (6) will be invertible. A smaller area, however, makes it more likely that the assumption of constant motion will hold. We can avoid balancing these two issues by replacing the assumption of constancy with a smoothness assumption [6]. That is, it is assumed that the model parameters $\boldsymbol{m}$ vary smoothly across space. A smoothness constraint on the contrast/brightness parameters has the added benefit of avoiding a degenerate solution where a pure intensity-based modulation is used to describe the mapping between images.

We begin with an error function, $E(\boldsymbol{m}) = E_1(\boldsymbol{m}) + E_2(\boldsymbol{m})$, that combines a smoothness constraint, $E_2(\boldsymbol{m})$, with the previous geometric and intensity transformation constraint, $E_1(\boldsymbol{m})$. The term $E_1(\boldsymbol{m})$ is defined as in Equation (8)

without the summation: $E_1(\boldsymbol{m}) = \left[k - \boldsymbol{c}^T \boldsymbol{m}\right]^2$, with k and $\boldsymbol{c}$ given by Equations (9) and (10). The new term $E_2(\boldsymbol{m})$ embodies the smoothness constraint:

$$E_2(\boldsymbol{m}) = \sum_{i=1}^{8} \lambda_i \left[\left(\frac{\partial m_i}{\partial x} \right)^2 + \left(\frac{\partial m_i}{\partial y} \right)^2 \right], \tag{11}$$

where λ_i is a positive constant that controls the relative weight given to the smoothness constraint on parameter m_i. This error term penalizes solutions proportional to the local change in each parameter across a small spatial neighborhood. In so doing, we allow for a locally smooth, but globally elastic transformation. The full error function $E(\boldsymbol{m})$ is minimized, as before, by differentiating, setting the result equal to zero and solving for $\boldsymbol{m}$. The derivative of $E_1(\boldsymbol{m})$ is $dE_1(\boldsymbol{m})/d\boldsymbol{m} = -2\boldsymbol{c}\left[k - \boldsymbol{c}^T \boldsymbol{m}\right]$. The derivative of $E_2(\boldsymbol{m})$ is computed by first expressing the partials, $\partial m_i/\partial x$ and $\partial m_i/\partial y$ with discrete approximations [6], and then differentiating, to yield $dE_2(\boldsymbol{m})/d\boldsymbol{m} = 2L(\overline{\boldsymbol{m}} - \boldsymbol{m})$, where $\overline{\boldsymbol{m}}$ is the component-wise average of $\boldsymbol{m}$ over a small spatial neighborhood, and L is an 8×8 diagonal matrix with diagonal elements λ_i, and zero off the diagonal. Setting $dE_1(\boldsymbol{m})/d\boldsymbol{m} + dE_2(\boldsymbol{m})/d\boldsymbol{m} = 0$, and solving for $\boldsymbol{m}$ at each pixel location yields an enormous linear system which is intractable to solve. Instead $\boldsymbol{m}$ is estimated in the following iterative manner [6]:

$$\boldsymbol{m}^{(j+1)} = \left(\boldsymbol{c}\,\boldsymbol{c}^T + L\right)^{-1} \left(\boldsymbol{c}\,k + L\overline{\boldsymbol{m}}^{(j)}\right). \tag{12}$$

The initial estimate $\boldsymbol{m}^{(0)}$ is determined from the closed-form solution of Section 2.2. On the $j + 1^{st}$ iteration $\overline{\boldsymbol{m}}^{(j)}$ is estimated from the previous estimate, $\boldsymbol{m}^{(j)}$.

The use of a smoothness constraint has the benefit that it yields a dense locally affine and smooth transformation. The drawback is that the minimization is no longer analytic. We have found, nevertheless, that the iterative minimization is quite stable and converges relatively quickly (see Section 2.5).

2.4 Partial Data

Inherent to the registration algorithm described above is the assumption that each region in the source image has a corresponding match in the target image. As illustrated in Fig. 2, this need not always be the case. Under such situations, our registration algorithm typically fails. One way to contend with partial or missing data is to employ a pre-processing segmentation step. We propose, however, a more unified approach in which the registration and segmentation are performed simultaneously.

We begin by assuming that each pixel in the source and target are either related through the intensity and geometric model of Equation (7), denoted as model M_1, or cannot be explained by this transformation and therefore belongs to an "outlier" model M_2. Pixels belonging to the outlier model are those that do not have a corresponding match between the source and target images. Assuming

that the pixels are spatially independent and identically distributed (iid), the likelihood of observing a pair of images is given by:

$$L(\boldsymbol{m}) = \prod_{x,y\in\Omega} P(\boldsymbol{q}(x,y)), \tag{13}$$

where, $\boldsymbol{q}(x,y)$ denotes the tuple of source, $m_7 f(x,y,t)+m_8$, and target, $f(m_1x+m_2y+m_5, m_3x+m_4y+m_6, t-1)$, image intensities, Equation (7). To simplify the optimization of the likelihood function, we consider the log-likelihood function:

$$\begin{aligned}\log[L(\boldsymbol{m})] &= \log\left[\prod_{x,y\in\Omega} P(\boldsymbol{q}(x,y))\right] \\ &= \sum_{x,y\in\Omega} \log\left[P(\boldsymbol{q}(x,y)|M_1)P(M_1) + P(\boldsymbol{q}(x,y)|M_2)P(M_2)\right]. \end{aligned}\tag{14}$$

Assuming that the priors on the models, $P(M_1)$ and $P(M_2)$, are equal, the log-likelihood function simplifies to:

$$\log[L(\boldsymbol{m})] = \sum_{x,y\in\Omega} \log\left[P(\boldsymbol{q}(x,y)|M_1) + P(\boldsymbol{q}(x,y)|M_2)\right], \tag{15}$$

where the factored additive constant is ignored for purposes of maximization. We assume next that the conditional probabilities take the following form:

$$\log[L(\boldsymbol{m})] = \sum_{x,y\in\Omega} \log\left[e^{-r^2(x,y)/\sigma^2} + e^{-c^2}\right]. \tag{16}$$

For model M_1 we assume a Gaussian distribution (with variance σ^2), where $r(x,y)$ is the residual error between the source and target defined as:

$$r(x,y) = [(m_7 f(x,y,t) + m_8) - (f(m_1x + m_2y + m_5, m_3x + m_4y + m_6, t-1))]. \tag{17}$$

For model M_2 we assume a uniform distribution (i.e., c is a constant). The log-likelihood function is maximized by differentiating, setting the result equal to zero and solving for $\boldsymbol{m}$:

$$\frac{d\log[L(\boldsymbol{m})]}{d\boldsymbol{m}} = \sum_{x,y\in\Omega} \frac{\frac{dr^2(x,y)}{d\boldsymbol{m}} e^{-r^2(x,y)/\sigma^2}}{e^{-r^2(x,y)/\sigma^2} + e^{-c^2}} = \sum_{x,y\in\Omega} \frac{dr^2(x,y)}{d\boldsymbol{m}} w(x,y) = 0, \tag{18}$$

where $w(\cdot)$ is defined to be the ratio of the exponential distributions. As in the previous sections, the residual $r(\cdot)$ is linearized with respect to the model parameters $\boldsymbol{m}$. The derivative of the residual, $dr^2(x,y)/d\boldsymbol{m}$, is then substituted into the above to yield:

$$\sum_{x,y\in\Omega} -2\boldsymbol{c}[k - \boldsymbol{c}^T\boldsymbol{m}]w = 0, \tag{19}$$

with $\boldsymbol{c}$ and k given by Equations (9) and (10), and, as before, all spatial parameters are dropped for notational convenience. Solving for the model parameters then yields the maximum likelihood estimator:

$$\boldsymbol{m} = \left[\sum_{x,y \in \Omega} (\boldsymbol{c}\boldsymbol{c}^T) w \right]^{-1} \left[\sum_{x,y \in \Omega} (\boldsymbol{c}k) w \right]. \tag{20}$$

Note that this solution is a weighted version of the earlier least-squares solution, Equation (6), where the weighting, w, is proportional to the likelihood that each pixel belongs to model M_1. As before, a smoothness constraint can be imposed to yield the following iterative estimator:

$$\boldsymbol{m}^{(j+1)} = \left((\boldsymbol{c}\ \boldsymbol{c}^T) w + L \right)^{-1} \left((\boldsymbol{c}\ k) w + L \overline{\boldsymbol{m}}^{(j)} \right). \tag{21}$$

This estimator for $\boldsymbol{m}$, however, requires an estimate of the weight w which itself requires an estimate of $\boldsymbol{m}$. The expectation/maximization algorithm (EM) [4] is used to resolve this circular estimator, and proceeds as follows:

1. E-step: compute the weights w (with an initial estimate of $\boldsymbol{m}$ from the solution of Section 2.3).
2. M-step: estimate the model parameters $\boldsymbol{m}$, Equation (21).
3. Repeat steps 1 and 2 until the difference between successive estimates of $\boldsymbol{m}$ is below a specified threshold.

The E-step is the segmentation stage, where pixels that do not have a corresponding match between source and target images have a close to zero weight w. These pixels are therefore given less consideration in the M-step which estimates the registration parameters $\boldsymbol{m}$. The EM algorithm allows for simultaneous segmentation and registration, and hence allows us to contend with missing data.

2.5 Implementation Details

While the formulation given in the previous sections is relatively straight-forward there are a number of implementation details that are critical for a successful implementation. First, in order to simplify the minimization, the error function of Equation (8) was derived through a Taylor-series expansion. A more accurate estimate of the actual error function can be determined using a Newton-Raphson style iterative scheme [11]. In particular, on each iteration, the estimated geometric transformation is applied to the source image, and a new transformation is estimated between the newly warped source and target image. As few as five iterations greatly improves the final estimate. Second, calculation of the spatial/temporal derivatives in Equations (9) and (10) is a crucial step. These derivatives are often computed using finite differences which typically yield poor approximations. We employ a set of derivative filters, specifically designed for multi-dimensional differentiation [5], that significantly improve the registration results. And third, a coarse-to-fine scheme is adopted in order to contend with

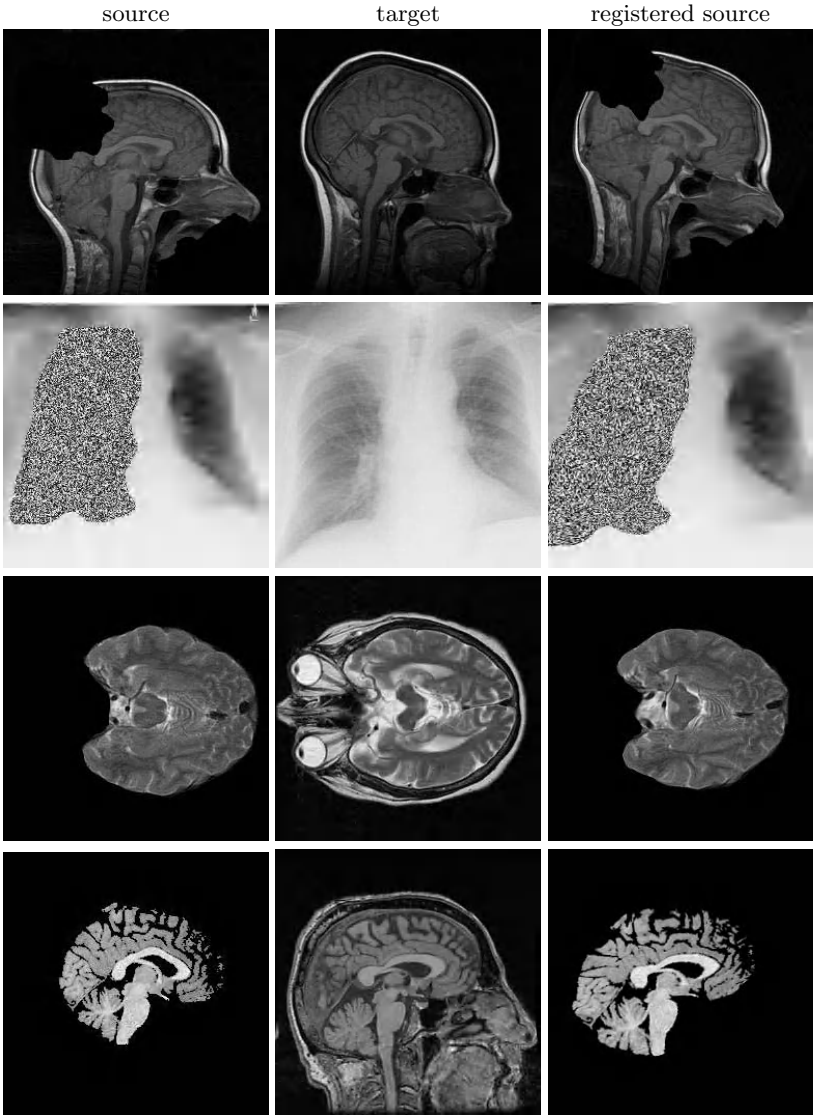


Fig. 2. Shown are synthetic and clinical examples of registration with significant portions of missing data.

larger motions [8,1]. A Gaussian pyramid is first built for both source and target images, and the full registration is estimated at the coarsest level. This estimate is used to warp the source image in the next level of the pyramid. A new estimate is computed at this level, and the process repeated throughout each level of the pyramid. The transformations at each level of the pyramid are accumulated yielding a single final transformation.

Fig. 3. Shown is an example of 3-D registration with partial data. The brighter regions shown with the registered source are the portions of the target that are missing in the source - these regions are superimposed to show the accuracy of the registration.

The generalization of the algorithm from 2-D images to 3-D volumes is relatively straight-forward. Briefly, to accommodate a 3-D affine transformation, an additional six affine parameters are added to the geometric and intensity transformation model of Equation (7). Linearization and minimization of this constraint proceeds as in the 2-D case. The smoothness constraint of Equation (11) takes on an additional $(\partial m_i/\partial z)^2$ term, and the iterative estimator of Equation (12) is of the same form, with k and $\boldsymbol{c}$ accommodating a different set of, now 3-D, spatial/temporal derivatives. The solution of Section (2.4) proceeds in a similar manner, with the initial constraint of Equation (13) updated to accommodate the 3-D geometric and intensity transformation model.

In the current MatLab implementation, running on a 2.8 GHz Linux machine, a pair of 256×256 images requires 4 minutes to register. A pair of $64 \times 64 \times 64$ volumes requires 30 minutes.

3 Results

We have tested the efficacy of our registration technique on both synthetic and clinical data in both 2-D and 3-D, Fig. 2-3. In the first two examples, regions were replaced with a uniform black or noise region. In the other examples the skull region was stripped from an axial, sagittal, and complete brain volume. In all cases, the registration is successful even with significant amounts of missing data. This registration would have failed without an explicit model of missing data incorporated directly into the registration algorithm. In all of the results of Fig. 1-3, all system parameters were held fixed.

4 Discussion

We have presented a general purpose registration algorithm. The geometric transformation is modelled as locally affine but globally smooth, and explicitly accounts for local and global variations in image intensities. An explicit

model of missing data is also incorporated, allowing us to simultaneously segment and register images with partial or missing data. All of the components are combined within an integrated framework yielding a robust and effective registration algorithm within and across different imaging modalities.

References

1. P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2(3):283–310, 1989.
2. R. Bansal, L. Staib, Z. Chen, A. Rangarajan, J. Knisely, R.Nath, and J.S. Duncan. A novel approach for the registration of 2D and 3D CT images for treatment setup verification in radiotherapy. In *Medical image computing and computer-assisted intervention (MICCAI)*, pages 1075–1086, 1998.
3. J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, 1994.
4. A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum lilelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, 99(1):1–38, 1977.
5. H. Farid and E.P. Simoncelli. Optimally rotation-equivariant directional derivative kernels. In *International Conference on Computer Analysis of Images and Patterns*, pages 207–214, Berlin, Germany, 1997.
6. B.K.P. Horn. *Robot Vision*. MIT Press, Cambridge, MA, 1986.
7. H. Lester and S.R. Arridge. A survey of hierarchical non-linear medical image registration. *Pattern Recognition*, 32(1):129–149, 1999.
8. B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence*, pages 674–679, Vancouver, 1981.
9. J.B.A. Maintz and M.A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.
10. S. Periaswamy and H. Farid. Elastic registration in the presence of intensity variations. *IEEE Transactions on Medical Imaging*, 2003 (in press).
11. J. Shi and C. Tomasi. Good features to track. In *Computer Vision and Pattern Recognition*, pages 593–600, Seatle, WA, USA, 1994.
12. S. Negahdaripour and C.-H. Yu. A generalized brightness change model for computing optical flow. In *International Conference of Computer Vision*, pages 2–11, Berlin, Germany, 1993.

Point Similarity Measure Based on Mutual Information

Peter Rogelj and Stanislav Kovačič

Faculty of Electrical Engineering
University of Ljubljana
Ljubljana, SI-1000, Slovenia
{peter.rogelj,stanislav.kovacic}@fe.uni-lj.si

Abstract. Registration of multi-modality images requires similarity measures that can deal with complex and unknown image intensity dependencies. Such measures have to rely on statistics, and consequently, they require relatively large image regions to operate. This makes the detection of localized image discrepancies difficult. As a solution we propose point similarity measures, which can measure similarity of arbitrarily small image regions, including similarity of individual image points. In this paper we present a point similarity measure derived from the mutual information. In addition to its extreme locality it can also avoid the interpolation artifacts and improve the spatial regularization to better suit the spatial deformation model.

1 Introduction

A criterion function used in registration procedures comprises some measure of image similarity, which is used for measuring the quality of image match, and possibly a geometric regularization, which prevents unrealistic deformations. However, not only the criterion function, also its implementation may have an important influence on the registration results. The influence of implementation is the most obvious for multi-modality similarity measures, when they are used to detect localized image discrepancies.

The ability to detect local image discrepancies is crucial for the success of non-rigid registration. The most straightforward approach to detect local image discrepancies is to measure the similarities of small image regions. In the case of conventional multi-modality similarity measures, e.g. mutual information measures [15,2,13], this is not directly applicable. Due to their statistical nature they require relatively large image regions to operate and thus cannot directly detect local image properties. Although these measures are not local, they are still locally sensitive, which enables assessment of local discrepancies by measuring global image similarity, i.e. similarity of the whole images, subject to local deformations. This is a general approach used by many authors, [12,8,11] to mention a few. The main problem with this approach is its high computational cost, which originates in large number of recomputations of global image similarity. This practically limits the dimensionality of the transformation and thus

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 112–121, 2003.

the locality of image discrepancies that can be corrected. In this work we focus on an alternative approach, which follows the basic idea of detecting local image discrepancies by measuring similarity of small image regions. To make this approach possible we introduced point similarity measures [10]. They are designed to measure similarity of arbitrarily small image regions, including similarity of individual image points, which also holds in the case of multi-modality data.

In this paper we present point similarity measure derived from the mutual information [15,2], and compare it with the original mutual information measure in terms of interpolation artifacts and spatial regularization.

2 Point Similarity Measures

We define point similarity measures as measures that can measure similarity of individual image points. Obviously they can also be used to measure similarity of image regions of arbitrary size. Point similarity measures can be derived from global similarity measures. Here we derive a point similarity from the mutual information, which is the most frequently used multi-modality similarity measure.

2.1 Point Similarity Measure Derived from the Mutual Information

Mutual information (MI) of two images (A and B) is defined by marginal and joint entropies:

$$MI = H(A) + H(B) - H(A, B), \tag{1}$$

and can be computed as follows:

$$MI = \sum_{\mathbf{i}} p(\mathbf{i}) \log \left(\frac{p\,(\mathbf{i})}{p\,(i_A)\,p\,(i_B)} \right). \tag{2}$$

Here, i_A and i_B are image intensities of images A and B, $\mathbf{i} = [i_A, i_B]$ denotes an intensity pair, $p(i_A)$ or $p(i_B)$ are marginal intensity distributions and $p(\mathbf{i})$ is a joint intensity distribution. Eq. (2) can be rewritten in the following form:

$$MI = \sum_{\mathbf{i}} \frac{N_{\mathbf{i}}}{N} \log \left(\frac{p\,(\mathbf{i})}{p\,(i_A)\,p\,(i_B)} \right) = \frac{1}{N} \sum_{v} \log \left(\frac{p(\mathbf{i}(v))}{p(i_A(v))p(i_B(v))} \right), \tag{3}$$

where $N_{\mathbf{i}}$ is the number of occurrences of intensity pair $\mathbf{i}$ and N is the total number of intensity pairs in the image, which equals the number of overlapping image voxels. Furthermore, $\mathbf{i}(v) = [i_A(v), i_B(v)]$ stands for image intensities located at voxel v. Note that the final summation is taken over the spatial image coordinates instead of the intensities. Thus, the global similarity MI can be treated as an average of point similarities $S_{MI}(v)$, defined for each voxel v.

$$MI = \frac{1}{N} \sum_{v} S_{MI}(v), \tag{4}$$

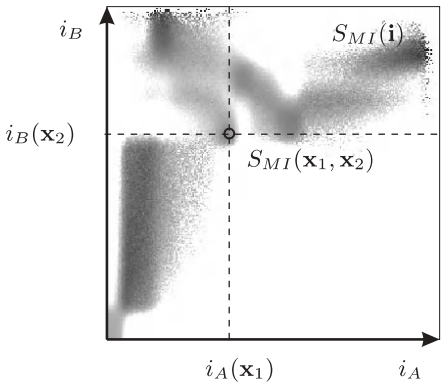


Fig. 1. Measurement of point similarity. The point similarity function $S_{MI}(\mathbf{i})$ is estimated from the whole images and defines the similarity with respect to the image intensities (darker color represents higher similarity). For example, a point similarity $S_{MI}(\mathbf{x}_1, \mathbf{x}_2)$ can be obtained from the point similarity function $S_{MI}(\mathbf{i})$ by pointing to a certain value by corresponding image intensity pair $\mathbf{i} = [i_A(\mathbf{x}_1), i_B(\mathbf{x}_2)]$.

$$S_{MI}(v) = \log\left(\frac{p(\mathbf{i}(v))}{p(i_A(v))p(i_B(v))}\right). \tag{5}$$

In general, the point similarity can be estimated not only for image voxels, but for each image point pair $[A(\mathbf{x}_1), B(\mathbf{x}_2)]$ from the corresponding intensity pair $\mathbf{i}(\mathbf{x}_1, \mathbf{x}_2) = [i_A(\mathbf{x}_1), i_B(\mathbf{x}_2)]$, such that $S_{MI}(\mathbf{x}_1, \mathbf{x}_2) = S_{MI}(\mathbf{i}(\mathbf{x}_1, \mathbf{x}_2))$, where

$$S_{MI}(\mathbf{i}) = \log\left(\frac{p(\mathbf{i})}{p(i_A)p(i_B)}\right). \tag{6}$$

The function $S_{MI}(\mathbf{i})$ is called a point similarity function and is an estimate of the intensity dependence between the images when they are correctly registered. The measurement of point similarity therefore consists of two steps. In the first step the point similarity function $S_{MI}(\mathbf{i})$ is estimated from the whole images A and B. In the second step point similarity $S_{MI}(\mathbf{x}_1, \mathbf{x}_2)$ is obtained from the point similarity function $S_{MI}(\mathbf{i})$ by simply pointing to a certain value by the corresponding image intensity pair $\mathbf{i}(\mathbf{x}_1, \mathbf{x}_2)$, as illustrated in Fig. 1. Note that the point similarity function needs to be estimated only once for all the measurements of point similarity at some image configuration, which makes the measurement of similarity computationally efficient.

2.2 Similarity of an Image Region

According to the Eq. (4) the mutual information of the whole images can be computed from point similarities by averaging. Let us rewrite it in the following form:

$$MI = \frac{1}{N}\sum_{v} S_{MI}(v) = \sum_{\mathbf{i}} \frac{N_{\mathbf{i}}}{N} S_{MI}(\mathbf{i}) = \sum_{\mathbf{i}} p(\mathbf{i}) S_{MI}(\mathbf{i}). \tag{7}$$

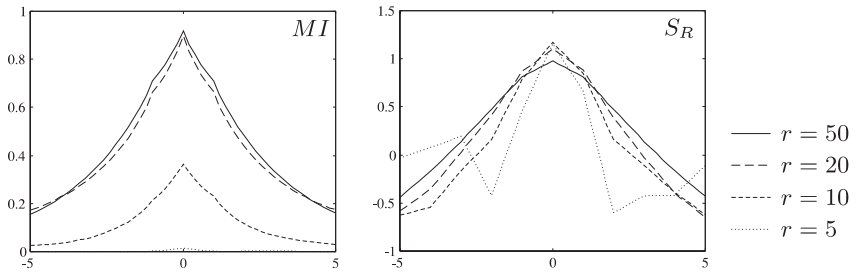


Fig. 2. Mutual information MI (left) and point based similarity measure S_R (right) with respect to image translation, for different sizes of image region r^3; $r = \{50, 20, 10, 5\}$. Point based similarity measure was always based on the same point similarity function $S_{MI}(\mathbf{i})$, obtained from the whole images at the correct image alignment. Note that the local sensitivity of the MI decreases with decreasing r, while the sensitivity of S_R remains practically the same.

The same principle could be used for computing similarity of smaller image regions, with the only difference that in this case the summation (averaging) runs only over the voxels in the region. Thus, the similarity S_R of the region R can be obtained by using probability distribution $p_R(\mathbf{i})$ of this region instead of probability distribution $p(\mathbf{i})$ of the whole images:

$$S_R = \frac{1}{N_R} \sum_{v \in R} S_{MI}(v) = \sum_{\mathbf{i}} p_R(\mathbf{i}) S_{MI}(\mathbf{i}), \tag{8}$$

where N_R is the number of voxels in the region R. Note that in all the cases the point similarity function $S_{MI}(\mathbf{i})$ is estimated from the whole images, in contrast to mutual information, which uses only the region that is being measured. This is a small but important difference between point based similarity measures and global measures, which enables the first ones to better assess the intensity dependence between the images and thus to improve the quality of local similarity measurement, see Fig. 2.

3 Similarity and Image Transformation

Image registration methods search for such a transformation that maximizes the image similarity. Let us analyze how the similarity changes when transforming image B with transformation $\mathbf{T}$. The transformation moves each point $B(\mathbf{x})$ from its original position $\mathbf{x}$ for some displacement $\mathbf{T}(\mathbf{x})$ to a new position $\mathbf{x} + \mathbf{T}(\mathbf{x})$, where it gets matched with a point $A(\mathbf{x} + \mathbf{T}(\mathbf{x}))$. One of the important issues is how the transformation changes the similarity of some small image region (or point) in case of subvoxel displacement $\mathbf{T}(\mathbf{x})$. A common phenomenon is the appearance of the interpolation artifacts, i.e. disproportionate change of similarity with respect to the transformation, which rules out the subvoxel accuracy. Some approaches that can reduce interpolation artifacts have been proposed [9,6] but can be applied only when the region size used for measuring the similarity is

large. In case of high-dimensional registration approaches, where gradient descent optimization method is usually used, the interpolation artifacts are even more problematic, because they can cause large image misregistration.

When observing a single point in image B, its point similarity $S_{MI}(A(\mathbf{x}+\mathbf{T}(\mathbf{x})), B(\mathbf{x}))$ may change due to two reasons. The first one is the change of point pair, and the second is a possible change of the intensity distributions, which changes the point similarity function $S(\mathbf{i})$, see Eq. (6).

Let us assume that the point similarity function $S_{MI}(\mathbf{i})$ does not change and that point similarities $S(A(\mathbf{x}+\mathbf{T}(\mathbf{x})), B(\mathbf{x}))$ change only because points in image B are compared with different points in image A. However, due to the discrete nature of the images and due to the image transformation, grid points in image B do not match exactly with grid points in image A, and measuring of point similarities requires interpolation. In case of mutual information there are two interpolation methods commonly used: interpolation of intensities and partial volume interpolation. Interpolation of intensities can also be employed in case of point similarities. However, the interpolation of intensities assumes a linear intensity dependence, which may not necessarily comply with the intensity dependence estimated from the images, and can cause interpolation artifacts, as shown in Fig. 3. To avoid the interpolation artifacts, we propose to interpolate similarities instead of intensities. This is related to partial volume interpolation, which may be used for estimation of global intensity distributions. Instead of interpolating the unknown intensity $i_A(\mathbf{x}+\mathbf{T}(\mathbf{x}))$ from intensities of neighboring voxels, we directly interpolate the point similarity $S(A(\mathbf{x}+\mathbf{T}(\mathbf{x})), B(\mathbf{x}))$ from similarities of point $B(\mathbf{x})$ to neighboring grid points in image A. The weights remain the same as in the case of interpolation of intensity. This approach results in a linear relationship between the point similarity and point displacement in a range of one image voxel, thereby avoiding the interpolation artifacts, see Fig. 3. The difference between results obtained by using different interpolation methods is illustrated in Fig. 4.

The point similarities could also change due to the change of the point similarity function. In general, the transformation $\mathbf{T}$ changes the marginal and joint intensity distributions $p(i_A)$, $p(i_B)$ and $p(\mathbf{i})$. Consequently, if the point similarity function $S_{MI}(\mathbf{i})$ is recomputed using the updated distributions as defined in Eq. (6), then it changes as well. The relation between the transformation and the change of intensity distributions is complex, nonlinear and depends on the information of the whole images. Furthermore, the relation between the intensity distributions and the corresponding point similarity function is not linear either (see Eq. (6)). Nevertheless, $S_{MI}(\mathbf{i})$ is always an approximation of the same intensity dependence of the images when they are correctly registered. The only difference between the obtained point similarity functions is in the quality of the estimation, which depends on the level of global image mismatch. When the point similarity function is estimated at better match, the similarity better distinguishes between correct matches and mismatches, while the positions of maxima that correspond to different tissue types in the point similarity function do not change, as shown in the experiment performed using simulated Brainweb

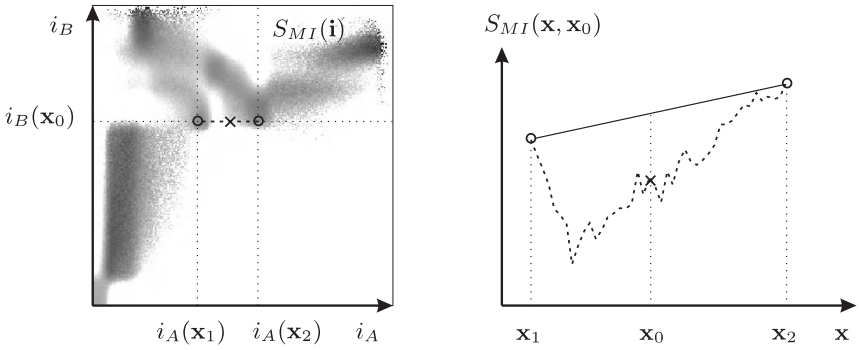


Fig. 3. Illustration of measuring point similarity with interpolation. Similarity between some voxel point $B(\mathbf{x}_0)$ and corresponding point $A(\mathbf{x}_0)$ requires interpolation. In case of interpolation of intensity, an intensity $i_A(\mathbf{x}_0)$ is interpolated from intensities of neighboring points $i_A(\mathbf{x}_1)$ and $i_A(\mathbf{x}_2)$, and the point similarity is $S(\mathbf{x}_0, \mathbf{x}_0) = S(i_A(\mathbf{x}_0), i_B(\mathbf{x}_0))$ (cross mark). Because the interpolated intensity does not comply with the complex intensity dependence estimated from the images, the similarity does not have a correct meaning, which introduces interpolation artifacts (dashed line). The problem can be solved by using interpolation of similarity, which interpolates the similarity $S(\mathbf{x}_0, \mathbf{x}_0)$ from point similarities $S(\mathbf{x}_1, \mathbf{x}_0)$ and $S(\mathbf{x}_2, \mathbf{x}_0)$, such that interpolation of intensity is not required and interpolation artifacts do not appear (solid line).

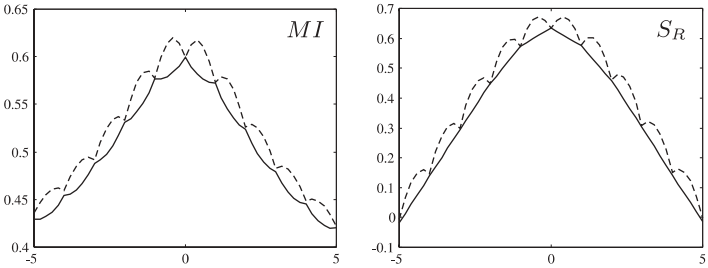


Fig. 4. An example of mutual information MI (left) and point based similarity S_R (right) with respect to image translation, for the two different interpolation methods. The dashed lines denote interpolation of intensity, while the solid lines denote partial volume interpolation for MI (left) and interpolation of similarity for S_R (right).

images [4] in Fig. 5. Consequently, the registration based on point similarity measures always tends towards the transformation that would be obtained when using point similarity function estimated from the registered images, see Fig. 6. Therefore, the complex nonlinear relation between the transformation and the similarity, which also reflects in interpolation artifacts, can be avoided by keeping the point similarity function fixed. However, when the point similarity function is obtained at large misalignment, the sensitivity of similarity measure is low. Therefore, to avoid the interpolation artifacts and still achieve good sensitivity, we propose to recompute the point similarity function only once per registration step or registration iteration.

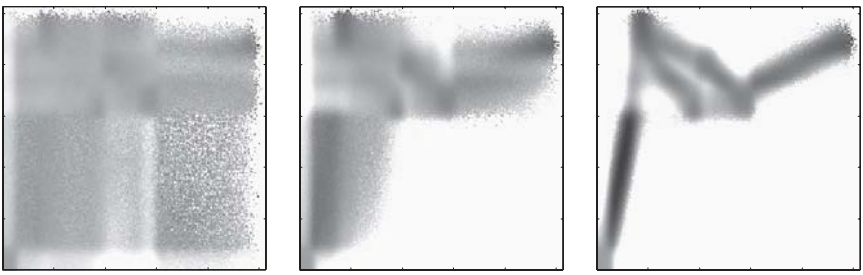


Fig. 5. Point similarity functions for simulated MRI-T1 and MRI-PD images of the head, at different levels of image mismatch: 10 mm displacement (left), 2 mm displacement (middle), and registered images (right). Darker color represents higher similarity. Note that the positions of maxima that correspond to different tissue types in point similarity function do not change.

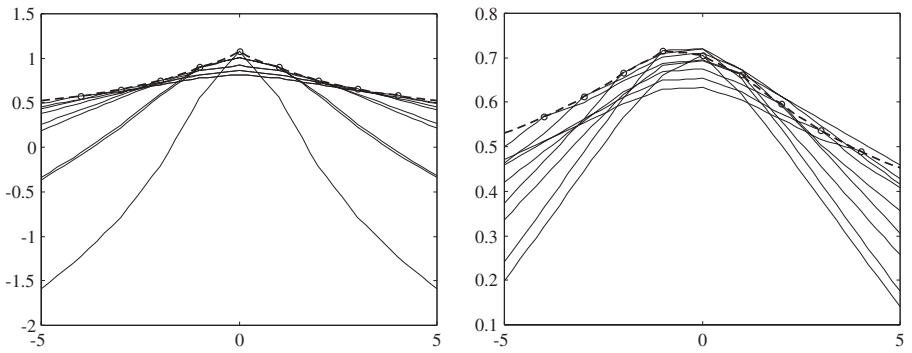


Fig. 6. Mutual information (dashed line) and point based similarities obtained using different estimations of point similarity function $S(\mathbf{i})$ (solid lines), with respect to image displacement. Similarity is measured between MRI-PD and MRI-T1 data, using simulated images (left) and real images of the head (right). Point similarity functions $S(\mathbf{i})$ were estimated at different image displacements. At this displacement the point based similarity equals the mutual information (marked with circles). All point based similarities reach the maximum at displacement 0, i.e. where images are correctly registered.

To summarize, to avoid the interpolation artifacts we keep the point similarity function fixed and use interpolation of similarity instead of interpolation of intensity.

4 Locality and Spatial Deformation Models

Point similarity measures push the limits of the locality into extreme. Consequently, similarity of one point does not presume any spatial relation with neighboring image points. However, as stated by some authors ([3,7]), matching of individual image points is ill-posed if they are matched independently. Reg-

istration with point similarity measures therefore requires regularization, which can be performed by a suitable spatial deformation model. Spatial deformation models are commonly used in high-dimensional registration, (for review of high-dimensional registration approaches see [14,5]). The model can follow physical properties of deformable materials (e.g. elasticity or viscosity) or simplified/fictious properties (e.g. Gaussian regularitzation). Nevertheless, the majority of such models can be performed by convolution filtering, as proposed by Bro-Nielsen [1]. For example, an incremental elastic registration can be performed iteratively as follows:

$$\mathbf{T}(\mathbf{x})^{(n+1)} = \mathbf{T}(\mathbf{x})^{(n)} + G_E * \frac{\partial S(\mathbf{x})}{\partial \mathbf{T}(\mathbf{x})}, \tag{9}$$

where G_E denotes a filter with impulse response of the elastic media, S stands for some similarity measure, and n is the iteration number.

Some regularization is also provided by similarity measures when they operate on larger image regions. The similarity of an image region can be obtained by averaging the point similarities, see Eq. (8). However, the averaging over a region surrounding a point $\mathbf{x}$ can also be performed by convolution filtering with some spatial filter G_R,

$$S_R(\mathbf{x}) = G_R * S_{MI}(\mathbf{x}). \tag{10}$$

Larger the region is, wider is the impulse response of the filter G_R and more global information is extracted from the point similarities. For example, when the region spreads over the whole images and the obtained similarity equals the global mutual information, only a global image properties, appropriate for global registration (e.g. rigid) are extracted. The averaging therefore represents a kind of regularization, which extracts more global knowledge from multiple more localized image features. The regularization caused by the similarity measures is substantial, such that when the region size used for measuring the similarity is large, the additional regularization with spatial deformation model may not be necessarily required, as in [7].

When the similarity of an image region is used in combination with a spatial deformation model, the Eq. (9) can be rewritten:

$$\mathbf{T}(\mathbf{x})^{(n+1)} = \mathbf{T}(\mathbf{x})^{(n)} + G_E * \frac{\partial G_R * S_P(\mathbf{x})}{\partial \mathbf{T}(\mathbf{x})} = \mathbf{T}(\mathbf{x})^{(n)} + G_E * G_R * \frac{\partial S_P(\mathbf{x})}{\partial \mathbf{T}(\mathbf{x})}. \tag{11}$$

Here, S_P is some point similarity measure, e.g. S_{MI}. Similar results could also be obtained for other high-dimensional registration methods. The regularization is therefore duplicated, which means that the final effect does not directly follow the spatial deformation model, see Fig. 7. Point similarity measures solve this problem. They are not regularized by G_R, such that regularization remains only in the domain of spatial deformation model, which gives a full control over the transformation properties.

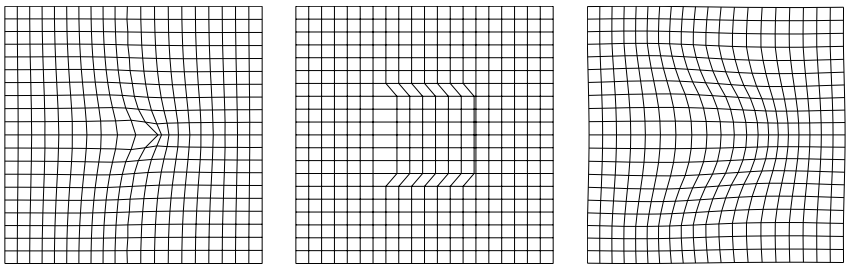


Fig. 7. An example of convolution filters used for regularizing non-rigid registration: elastic filter G_E (left), filter G_R that corresponds to region averaging (middle), and their convolution $G_E * G_R$ (right).

5 Conclusion

In this paper we presented a point similarity measure derived from the mutual information. This measure can directly detect localized image discrepancies, by measuring similarity of small image regions or individual image points. It is computationally efficient, as the point similarity function is computed only once for a given image configuration. The other advantage is the ability to avoid the interpolation artifacts. That is because point similarity measures may be estimated using the same estimation of image intensity dependencies, regardless of the actual image transformation. This could also be advantageous in the case of rigid registration. Finally, point similarity measures do not constrain the transformation, which is especially important when performing a high-dimensional registration. In this case the regularization remains only in the domain of a spatial deformation model, which does not interfere with the similarity measures.

References

1. M. Bro-Nielsen. *Medical Image Registration and Surgery Simulation.* PhD thesis, Department of Mathematical Modelling, Technical University of Denmark, 1996.
2. A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multi-modality image registration based on information theory. In Y. Bizais, C. Barillot, and R. Di Paola, editors, *Information processing in medical imaging 1995*, pages 263–274. Kluwer Academic, 1995.
3. T. Gaens, F. Maes, D. Vandermeulen, and Suetens P. Non-rigid multimodal image registration using mutual information. In W.M. Wells, A. Colchester, and S. Delp, editors, *Proceedings of the 1st International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI'98*, number 1496 in Lecture Notes in Computer Science, pages 1099–1106, MIT, Cambridge, MA, USA, October 1998. Springer-Verlag.
4. R.K.-S. Kwan, A.C. Evans, and G. B. Pike. An extensible MRI simulator for post-processing evaluation. In *Visualization in Biomedical Computing (VBC'96)*, volume 1131 of *Lecture Notes in Computer Science*, pages 135–140. Springer-Verlag, May 1996.

5. H. Lester and S. R. Arridge. A survey of hierarchical non-linear medical image registration. *Pattern Recognition*, 32(1):129–149, 1999.
6. B. Likar and F. Pernuš. A hierarchical approach to elastic registration based on mutual information. *Image and Vision Computing*, 19:33–44, 2001.
7. J. B. A. Maintz, H. W. Meijering, and M. A Viergever. General multimodal elastic registration based on mutual information. In K.M. Hanson, editor, *Medical Imaging 1998: Image Processing*, volume 3338 of *Proc. SPIE*, pages 144–154. SPIE Press, Bellingham, WA, 1998.
8. D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank. Non-rigid multimodality image registration. In M. Sonka and K.M. Hanson, editors, *Medical Imaging 2001: Image Processing*, volume 4322 of *Proc. SPIE*. SPIE Press, Bellingham, WA, 2001.
9. J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Interpolation artefacts in mutual information based image registration. *Computer Vision and Image Understanding*, 77(2):211–232, 2000.
10. P. Rogelj and S. Kovačič. Similarity measures for non-rigid registration. In M. Sonka and K.M. Hanson, editors, *Medical Imaging 2001: Image Processing*, volume 4322 of *Proc. SPIE*, pages 569–578. SPIE Press, Bellingham, WA, 2001.
11. G. K. Rohde, A Aldroubi, and B. M. Dawant. Adaptive free-form deformation for inter-patient medical image registration. In *The Proceedings of the SPIE Symposium on Medical Imaging 2001*, 2001.
12. D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach, and D.J. Hawkes. Nonrigid registration using free-form deformations: Application to breast mr images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, August 1999.
13. C. Studholme, D.L.G. Hill, and D.J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32:71–86, 1999.
14. P. Thompson and A. W. Toga. Warping strategies for intersubject registration. In I. Bankman, editor, *Handbook of Medical Image Processing*. Academic Press, 1999.
15. P. Viola and W. Wells III. Alignment by maximization of mutual information. In *Proceedings of the 5th International Conference on Computer Vision*, pages 16–23, 1995.

An Information Theoretic Approach for Non-rigid Image Registration Using Voxel Class Probabilities

E. D'Agostino, F. Maes*, D. Vandermeulen, and P. Suetens

Katholieke Universiteit Leuven
Faculties of Medicine and Engineering
Medical Image Computing (Radiology - ESAT/PSI)
University Hospital Gasthuisberg
Herestraat 49, B-3000 Leuven, Belgium
Emiliano.DAgostino@uz.kuleuven.ac.be

Abstract. We propose a multimodal free-form registration algorithm that matches voxel class labels rather than image intensities. Individual voxels are displaced such as to minimize the Kullback-Leibler distance between the actual and ideal joint probability distribution of voxel class labels, which are assigned to each image individually by a previous segmentation process. We evaluate the performance of the method for inter-subject brain registration with simulated deformations, using a viscous fluid model for regularization. The root mean square difference between recovered and ground truth deformations is smaller than 1 voxel.

1 Introduction

While maximization of mutual information (MMI) of corresponding voxel intensities [4] has been demonstrated to be highly successful for affine image registration of multimodal medical images in a variety of different applications, extending the MMI approach to non-rigid image registration is still an active area of research. While changes in the affine registration parameters affect the registration over the entire image domain and therefore in general have substantial impact on the joint intensity histogram and on the MI criterion itself, local non-rigid deformation changes typically result in only subtle histogram changes that are more difficult to assess using MI. Various approaches have been presented that differ in the way the criterion is evaluated for local non-rigid changes in the registration transformation and in the regularization that is applied to exclude non-realistic deformations and to make the non-rigid registration problem well-posed. Rueckert *et al.* [8] used a B-spline representation of the deformation field to impose local smoothness, whose parameters are optimized iteratively such that MI measured globally over the entire image domain is maximized. Hermosillo *et al.* [2] and D'Agostino *et al.* [1] derived the gradient of MI with

* Frederik Maes is Postdoctoral Fellow of the Fund for Scientific Research - Flanders (FWO-Vlaanderen, Belgium).

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 122–131, 2003.

respect to individual voxel displacements using a continuous and differentiable joint histogram constructed by Parzen windowing, yielding a voxel-wise force field that was used to drive free-form deformation regularized by elastic [2] or viscous fluid [1] deformation models.

However, all these approaches have so far considered MI of corresponding voxel intensities, implicitly assuming that these are related to objects that are to be aligned by registration. In fact, in the MMI criterion formulated by Maes *et al.* [4], image segmentation and labelling is implicit by the intensity rescaling and binning that is applied when contructing the joint intensity histogram. In this paper, we explore the possibility of non-rigid image registration by maximizing an information theoretic measure of the similarity of voxel object labels directly, rather than of voxel intensities. A related approach was presented by Rohlfing *et al.* [7], using nearest neighbour interpolation of integer-valued label images obtained by manual object delineation in each of the images to be registered. In our application, which is inter-subject MR brain image matching, such labels are obtained by intensity-based tissue segmentation, assigning each voxel a probability to belong to a particular tissue class, rather than a single label value as in [7], and using partial volume interpolation [4] to construct the joint label histogram. The advantage of such object-based approach is two-fold. Firstly, the ideal correspondence of object labels in both images is known, which is not the case for the multimodal intensities themselves, which is exploited in the registration criterion as explained in Section 2. Secondly, image segmentation allows to indirectly incorporate complex object intensity models and spatial context in the registration process, which may help to guide and constrain the non-rigid matching in regions where image information by itself is ambiguous. Our long term goal is to merge image segmentation and non-rigid registration is a single framework, whereby the result of each is iteratively improved by the output of the other.

2 Method

2.1 Similarity Measure

When two monomodal images are perfectly aligned, their joint histogram becomes diagonal, such that the quality of the registration can be evaluated by off-diagonal histogram dispersion. For multimodal images however, the joint histogram at registration depends on the nature of the data itself and can in general not be predicted in advance. However, by first segmenting each image separately by assigning each voxel an object label, the multimodal image registration problem can be reduced to a monomodal one that consists of aligning identically labelled voxels (subject to appropriate regularization constraints as described below) rather than the multimodal intensities of the original images. If a unique object label is assigned to each voxel in each image (and only a small number of objects is considered), non-rigid registration of the resulting crisp label images is likely to be hampered by the lack of registration features in homogeneously labeled regions and by interpolation artefacts at object boundaries.

Instead, for the brain MR images considered in this paper, we create label images by intensity-based pixel classification using the fully automated model-based algorithm by Van Leemput *et al.* [10]. Each voxel i is assigned a probability $c_{i,k}$ to belong to one of four tissue classes k (gray matter, white matter, CSF, and other), such that for each voxel $\sum_{k=1}^{4} c_{i,k} = 1$. A 4×4 joint class distribution p can then be constructed assuming that the probability of corresponding voxels in the two images to belong to a particular class are independent:

$$p_{\text{actual}} = \sum_{(i,j)} P_{ij}^{T,R} = \sum_{(i,j)} \begin{vmatrix} c_{i,1}^T.c_{j,1}^R & c_{i,1}^T.c_{j,2}^R & c_{i,1}^T.c_{j,3}^R & c_{i,1}^T.c_{j,4}^R \\ c_{i,2}^T.c_{j,1}^R & c_{i,2}^T.c_{j,2}^R & c_{i,2}^T.c_{j,3}^R & c_{i,2}^T.c_{j,4}^R \\ c_{i,3}^T.c_{j,1}^R & c_{i,3}^T.c_{j,2}^R & c_{i,3}^T.c_{j,3}^R & c_{i,3}^T.c_{j,4}^R \\ c_{i,4}^T.c_{j,1}^R & c_{i,4}^T.c_{j,2}^R & c_{i,4}^T.c_{j,3}^R & c_{i,4}^T.c_{j,4}^R \end{vmatrix} \tag{1}$$

with $c_{i,k}^T$ and $c_{j,k}^R$ the class probabilities of the pairs of corresponding voxels i and j in the template image T and reference image R respectively. When the images T and R are properly aligned, the matrix p is expected to be diagonally dominant. By considering the ideal case whereby tissue labels in T and R are identical and T has been perfectly warped onto R, a joint class probability reference model can be constructed as

$$p_{\text{ideal}} = \sum_{(i,j)} P_{ii}^{R,R} = \sum_{(i,j)} \begin{vmatrix} c_{i,1}^R.c_{i,1}^R & c_{i,1}^R.c_{i,2}^R & c_{i,1}^R.c_{i,3}^R & c_{i,1}^R.c_{i,4}^R \\ c_{i,2}^R.c_{i,1}^R & c_{i,2}^R.c_{i,2}^R & c_{i,2}^R.c_{i,3}^R & c_{i,2}^R.c_{i,4}^R \\ c_{i,3}^R.c_{i,1}^R & c_{i,3}^R.c_{i,2}^R & c_{i,3}^R.c_{i,3}^R & c_{i,3}^R.c_{i,4}^R \\ c_{i,4}^R.c_{i,1}^R & c_{i,4}^R.c_{i,2}^R & c_{i,4}^R.c_{i,3}^R & c_{i,4}^R.c_{i,4}^R \end{vmatrix} \tag{2}$$

We propose to use the Kullback-Leibler distance $D(p_{\text{ideal}}||p_{\text{actual}})$ between the ideal (fixed) and the actual (registration dependent) joint class probability distribution as registration measure, which needs to be minimized by optimization of the registration parameters:

$$D(p_{\text{ideal}}||p_{\text{actual}}) = \sum_{k} p_{\text{ideal}}.\log \frac{p_{\text{ideal}}}{p_{\text{actual}}} \tag{3}$$

with the sum taken over all 4×4 entries in p_{ideal}.

In practice, we construct p by transforming voxels from R into T, using partial volume (PV) interpolation [4] in the space of voxel labels to distribute the contribution of each sample i in R to the joint histogram over its 8 (in 3-D) nearest neighbours j on the grid of T without need for interpolation of class labels or probabilities directly:

$$p_{\text{real}} = \sum_{i} \sum_{j} w_{ij} P_{ij}^{T,R} \tag{4}$$

with w_{ij} the trilinear interpolation weights and $P_{ij}^{T,R}$ as defined above. The weights w_{ij} are continuous and differentiable functions of the transformation $\boldsymbol{u}(\boldsymbol{x})$ from $\boldsymbol{x}$ in image R onto $\boldsymbol{x} - \boldsymbol{u}(\boldsymbol{x})$ in image T [5].

2.2 Force Field Computation

The effect of a local displacement of a single voxel i along each of the three coordinate axes on the registration criterion $D(p_{\text{ideal}}||p_{\text{actual}})$ can be computed by derivation of D with respect to the displacement $\boldsymbol{u}_i = \boldsymbol{u}(\boldsymbol{x}_i)$, using a similar approach as in [5] and assuming that the number of voxels within the region of overlap of both images is constant:

$$\frac{\partial D}{\partial \boldsymbol{u}_i} = \sum_k \frac{\partial D}{\partial p_{\text{actual}}} . \frac{\partial p_{\text{actual}}}{\partial \boldsymbol{u}_i} = -\sum_k \frac{p_{\text{ideal}}}{p_{\text{actual}}} . \frac{\partial p_{\text{actual}}}{\partial \boldsymbol{u}_i} \tag{5}$$

$$\frac{\partial p_{\text{actual}}}{\partial \boldsymbol{u}_i} = \sum_j \frac{\partial w_{ij}}{\partial \boldsymbol{u}_i} . P_{ij}^{T,R} \tag{6}$$

Hence, a voxel-wise force field $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{u})$ can be defined that tends to displace each voxel such as to minimize D:

$$\boldsymbol{F}(\boldsymbol{x}_i, \boldsymbol{u}) = -\frac{\partial D}{\partial \boldsymbol{u}_i} = \sum_k \frac{p_{\text{ideal}}}{p_{\text{actual}}} \left(\sum_j \frac{\partial w_{ij}}{\partial \boldsymbol{u}_i} . P_{ij}^{T,R} \right) \tag{7}$$

2.3 Viscous Fluid Regularization

To evaluate the merits of the new registration criterion, we adopt the free-form registration approach of [1] and use the force field $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{u})$ to drive a viscous fluid regularizer by iteratively solving its Navier-Stokes governing equation:

$$\nabla^2 \boldsymbol{v} + \nabla \left(\nabla . \boldsymbol{v} \right) + \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{u}) = 0 \tag{8}$$

with $\boldsymbol{v}(\boldsymbol{x}, t)$ the deformation velocity experienced by a particle at position $\boldsymbol{x}$. An approximate solution of (8) is obtained by convolution with a Gaussian kernel:

$$\boldsymbol{v} = \psi \star \boldsymbol{F} \tag{9}$$

and the deformation field $\boldsymbol{u}^{(k+1)}$ at iteration $(k+1)$ is found by integration over time:

$$\boldsymbol{R}^{(k)} = \boldsymbol{v}^{(k)} - \sum_{i=1}^{3} v_i^{(k)} \left[\frac{\partial \boldsymbol{u}^{(k)}}{\partial x_i} \right] \tag{10}$$

$$\boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} + \boldsymbol{R}^{(k)} . \Delta t \tag{11}$$

The time step Δt is constrained by $\Delta t \leq \max(\|\boldsymbol{R}\|) . \Delta u$, with Δu the maximal voxel displacement that is allowed in one iteration. Regridding and template propagation are used as in [1] to preserve topology.

3 Results

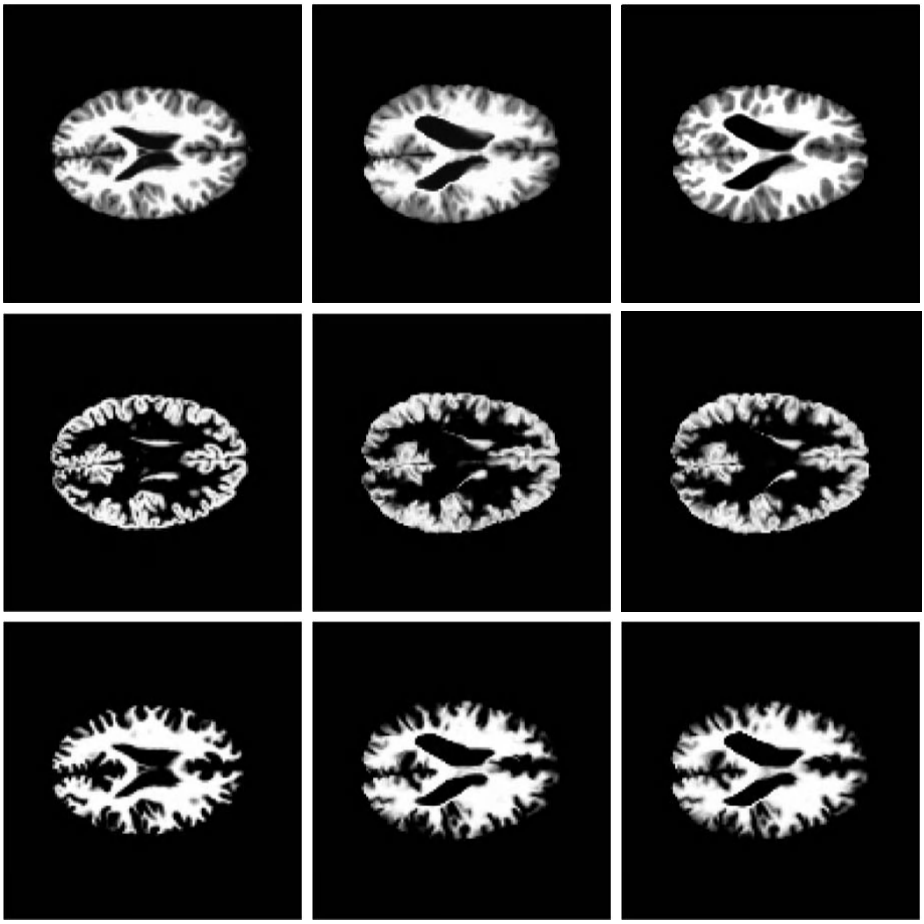
The method described above was implemented in Matlab, with the image resampling and histogram computation coded in C. The maximal voxel displacement Δu at each iteration was set to 0.25 voxels and regridding was performed when the Jacobian of the deformation field became smaller than 0.5. Iterations were continued as long as the Kullback-Leibler distance D decreased, with a maximal number of iterations of 180. Computation time for matching two images of size 128x128x80 is about 7 minutes on a Pentium IV 2.2 GHz PC.

We evaluated the performance of the method and compared it with the voxel intensity based MMI method described in [1] using the same validation strategy as in [1]. The method was validated on simulated images of size 128x128x80 generated by the BrainWeb MR simulator [9], which were non-rigidly deformed by known deformation fields $\boldsymbol{u}^*$. These were generated by matching the T1-weighted BrainWeb image to real T1-weighted images of 3 periventricular leukomalacia patients, typically showing enlarged ventricles. The ground truth deformation $\boldsymbol{u}^*$ was generated twice, once using the voxel intensity based (VI) MMI method of [1] and once using the label probability based (LP) method described in this paper. We evaluate how well the recovered deformation $\boldsymbol{u}$, obtained by matching the original T1-weighted BrainWeb image to its deformed version using either of both methods, resembles the ground truth $\boldsymbol{u}^*$. Both deformations were compared by their root mean square (RMS) error Δu evaluated in voxel units over all brain voxels $\mathcal{B}$:

$$\Delta T = \sqrt{\frac{1}{N_{\mathcal{B}}} \sum_{\mathcal{B}} (|\boldsymbol{u}(\boldsymbol{x}) - \boldsymbol{u}^*(\boldsymbol{x})|)^2} \tag{12}$$

Error values for 3 different cases are summarized in table 1. Although the RMS error is subvoxel small in all cases, the method that generated the ground truth deformation performs generally better. A typical registration result obtained with the method proposed in this paper is illustrated in figure 1, showing the template, target and matched template images and the corresponding WM and GM maps. Deformation fields obtained with both methods are shown in figure 2.

Figure 3 displays the error $|\boldsymbol{u}(\boldsymbol{x}) - \boldsymbol{u}^*(\boldsymbol{x})|$ over a particular axial slice of the deformation field $\boldsymbol{u}$ obtained with either of both methods compared with the ground truth $\boldsymbol{u}^*$ computed with the voxel intensity based MMI method of [1]. The norm $|\boldsymbol{u}^*|$ of the deformations to be recovered within the brain region in this slice has a mean of 0.86 voxels, standard deviation of 0.87 voxels and maximum of 4.61 voxels. Average (maximal) errors are 0.25 (1.69) and 0.49 (2.84) voxels for the VI and LP based methods respectively. The difference between VI and LP registration solutions (figure 3(d)) is largest near the high contrast boundaries of the brain and ventricles (mean difference is 0.45 voxels, maximum is 2.50 voxels).



Initial:

.1669	.0926	.0131	.0031
.0509	.2512	.0606	.0109
.0070	.0546	.0788	.0135
.0039	.0296	.0434	.1198

Final:

.1579	.0634	.0045	.0021
.0637	.2908	.0589	.0092
.0049	.0619	.1076	.0196
.0022	.0121	.0249	.1164

Model:

.1586	.0636	.0044	.0021
.0636	.2937	.0603	.0105
.0044	.0603	.1093	.0218
.0021	.0105	.0218	.1129

Final - Initial:

-.0090	-.0292	-.0086	-.0010
.0128	.0395	-.0017	-.0017
-.0021	.0073	.0288	.0060
-.0017	-.0175	-.0185	-.0034

Fig. 1. Original intensity image and GM and WM maps for the template image (left), template image warped to reference image (middle) and reference image (right). The joint class distributions capture the coincidence of WM, GM, CSF and OTHER class voxels in the reference (columns) and template (rows) images.

Table 1. Root mean square error Δu in voxels between ground thruth (u^*) and recovered (u) deformation fields for 3 different cases, computed either with the label probability (PB) criterion presented here or with the voxel intensity (VI) based MMI approach of [1].

u^*	VI		LP	
u	VI	LP	VI	LP
Case 1	0.42	0.53	0.54	0.38
Case 2	0.46	0.56	0.45	0.35
Case 3	0.47	0.46	0.43	0.33

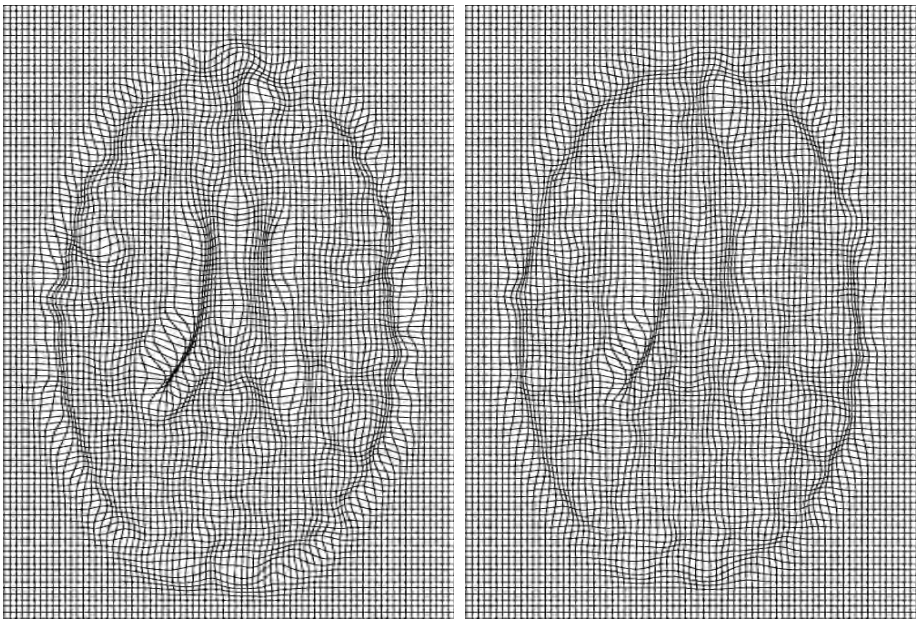


Fig. 2. Left: Deformation recovered by the MMI based method of [1]; right: deformation recovered by method proposed here.

4 Discussion

We present an algorithm for non-rigid image registration that optimizes an information theoretic measure of voxel class probabilities rather than voxel intensities directly. Class probabilities are assigned to each voxel by intensity-based tissue classification of each image individually and the joint class distribution is obtained by assuming that corresponding voxel class labels in each image are independent. The segmentation introduces correspondence information about the voxel features in each image, which is not available with intensity-based multimodality image registration. Hence, a good model for the joint class distribution at registration can be estimated by considering the case of perfect alignment of an image with itself. A related approach, in the sense that it also used a prior joint distribution model, but involving intensity values rather than class

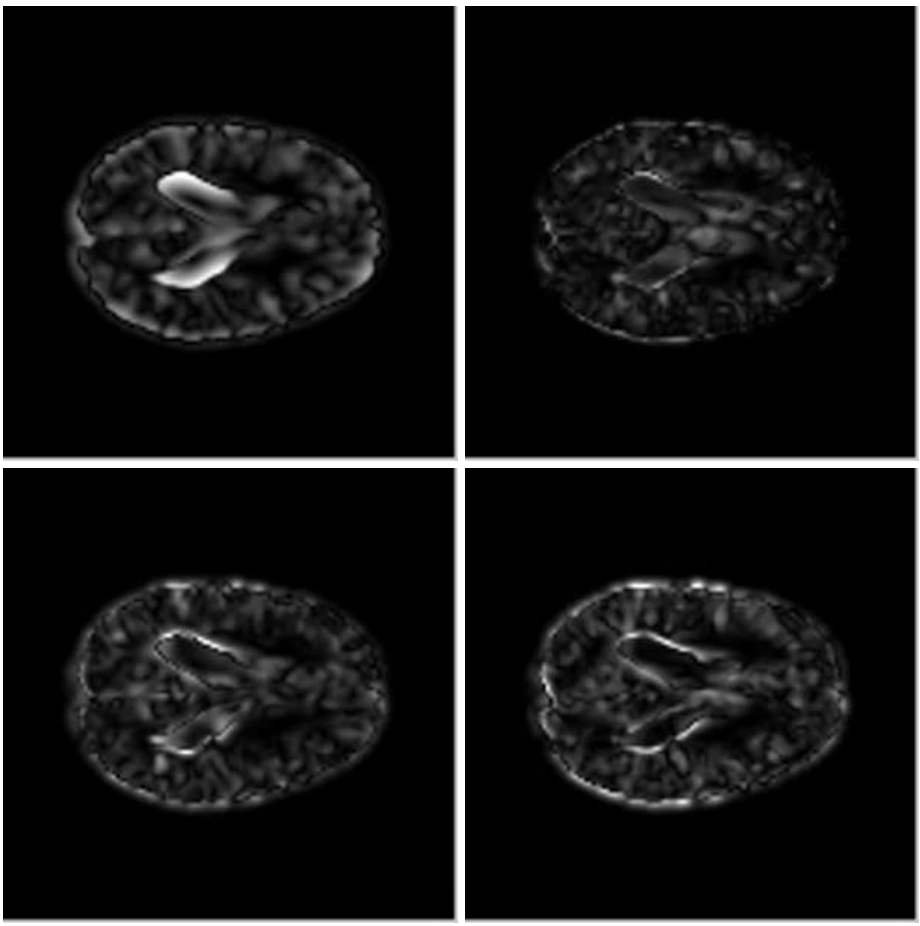


Fig. 3. (a) Norm of the ground truth displacement $|\boldsymbol{u}^*|$ to be recovered within this slice. The deformation is largest at the ventricles and at the brain contour; (b,c) Error norm $|\boldsymbol{u}^* - \boldsymbol{u}|$ for the VI and LP based approaches respectively; (d) Difference $|\boldsymbol{u}^{VI} - \boldsymbol{u}^{LP}|$ between the VI and LP registration solutions. The corresponding deformation fields are shown in figure 2.

labels and derived from training images, was presented by [3]. We propose to use the Kullback-Leibler distance between the observed and the model class distribution as a new registration measure. Inspection of the initial and final actual histograms p_{actual} and of the model histogram p_{model} depicted in figure 1, demonstrates that the joint class histogram after non-rigid matching is indeed much more similar to the model than before registration. The change in joint class probabilities prior and after registration shows that a clustering has occurred along each column, such that after registration template class labels corresponding to a particular reference class label are distributed similarly as in the model distribution.

The joint class distribution p_{actual} is estimated during registration using PV interpolation [4] such that it varies smoothly with individual voxel displacements and can be analytically differentiated. A force field is thus obtained that acts to displace individual voxels such that the Kullback-Leibler distance between the observed and the model distribution is minimized. In the approach presented in [1], the gradient of MI is computed with respect to individual voxel displacements by modelling the joint intensity histogram as a continuous and differentiable function of the image intensity itself using a Parzen estimator. The approach presented here however is completely discrete due to the PV interpolation scheme. Bias in the registration criterion by grid-aligning displacements as reported for MMI [6] is not an issue here, due to the fuzzy nature of the class labels. In contrast with the method of [1], the force field (7) does not depend on the image intensity gradient of the template image, which may explain the larger smoothness of the resulting deformation field as observed in figure 2. The observation that the deformations obtained with both approaches differ the most near high contrast object boundaries, as illustrated in figure 3(d), can be understood by the fact that the weighting of the force field with the image intensity gradient in the method of [1] will have its highest impact precisely at such locations.

We demonstrated how the force field can be applied to drive free-form non-rigid registration using a viscous fluid regularization model. However, while the viscous fluid model is appropriate for inter-subject brain registration as focussed on here (whereby large deformations have to be recovered in patients showing enlarged ventricles), other regularization schemes (e.g. elastic [2]) could be used as well. Validating the performance of various regularization schemes for specific non-rigid registration applications is an important research topic and outside the scope of this paper.

A main advantage of this approach is that a priori model knowledge regarding intensity appearance or spatial context of the objects in each image can be incorporated by a model-based segmentation strategy. In our case for instance, focussing on inter-subject brain image registration, WM, GM and CSF are segmented assuming a Gaussian mixture intensity model, while also accounting for MR intensity inhomogeneity and incorporating a priori information of the expected spatial distribution of tissues in the brain [10]. Using class labels as features for non-rigid image registration opens perspectives for integrating registration and segmentation as two cooperative processes in a single framework, for instance by considering one of the images as an atlas that is non-rigidly warped onto the other and that provides a priori tissue distribution maps to guide the segmentation of the other image. Future work will be directed towards this goal.

References

1. E. D'Agostino, F. Maes, D. Vandermeulen, and P. Suetens. A viscous fluid model for multimodal non-rigid image registration using mutual information. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI'02)*, volume 2488 of *Lecture Notes in Computer Science*, pages 541–548, Tokyo, Japan, September 2002. Springer-Verlag, Berlin.

2. G. Hermosillo, C. Chef d'Hotel, and O. Faugeras. A variational approach to multi-modal image matching. Technical Report 4117, INRIA-ROBOTVIS, Sophia Antipolis, France, February 2001.
3. M.E. Leventon and W.E.L. Grimson. Multi-modal volume registration using joint intensity distribution. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI'98)*, volume 1496 of *Lecture Notes in Computer Science*, Cambridge, Massachusetts, October 1998. Springer-Verlag, Berlin.
4. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, April 1997.
5. F. Maes, D. Vandermeulen, and P. Suetens. Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information. *Medical Image Analysis*, 3(4):373–386, 1999.
6. J.P.W. Pluim, J.B.A. Maintz, and M.A. Viergever. Interpolation artefacts in mutual information-based image registration. *Computer Vision and Image Understanding*, 77(2):211–232, 2000.
7. T. Rohlfing, R. Brandtand C.R. Maurer Jr., and R. Menzel. Bee brains, b-splines and computational democracy: Generating an average shape atlas. In *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA'01)*, volume 1496 of *Lecture Notes in Computer Science*, pages 187–194, Kauai, Hawaii, December 2001. IEEE Computer Society.
8. D. Rueckert, L.I. Sonoda, C. Hayes, D. Hill, M.O. Leach, and D.J. Hawkes. Non-rigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.
9. BrainWeb MR simulator. Available at http://www.bic.mni.mcgill.ca/brainweb/.
10. K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automated model-based tissue classification of MR images of the brain. *IEEE Transactions on Medical Imaging*, 18(10):897–908, 1999.

Spatial Information in Entropy-Based Image Registration

Mert R. Sabuncu and Peter J. Ramadge

Department of Electrical Engineering, Princeton University
Princeton, NJ 08544

Abstract. Information-theoretic approaches (e.g. mutual information) have yielded accurate and robust similarity measures for multi-modal and mono-modal image registration. However, recent research suggests that registration based on mutual information has room for improvement. The paper proposes a method for including spatial information in this approach by using spatial feature vectors obtained from the images. A minimum spanning tree algorithm is employed to compute an estimate of the conditional entropy in higher dimensions. The paper includes the theory to motivate the proposed similarity measure. Experimental results indicate that the suggested method can achieve a more robust and accurate registration compared to similarity measures that don't include spatial information.

1 Introduction

The goal of image registration is to geometrically transform one image so that physical (for medical imaging, physiological) correspondences line up. This is usually a precursor to comparison or fusion of the images. The problem is multi-modal if images are obtained through two different imaging devices, e.g. functional Magnetic Resonance images (fMRI) and structural MR images. A popular approach to this problem defines a similarity measure between the images and performs an optimization over allowed transformations to maximize this measure. Here, we consider "intrinsic" image registration methods that are based on voxel or pixel intensities. In the literature, many different similarity measures have been proposed under this category. However, most of these methods assume a strict prior distribution on the noise (such as i.i.d. Gaussian noise) and/or a linear relation between the two image intensities. Due to these assumptions, these methods generally perform poorly for multi-modal registration.

An information-theoretic approach was recently introduced by Viola et al. [11]. The main idea is to use a distance measure based on Mutual Information (MI) and its variants [8]. Using MI was motivated by a maximum likelihood (ML) solution to a certain model. However, the assumption that the imaging function doesn't depend on spatial position [12] is very restrictive. Many researchers [7,9] noticed this flaw and proposed heuristic improvements. One approach is to use higher dimensional feature vectors to capture the spatial information. However, until recently a good estimate for the entropy of higher dimensional p.d.f.'s was not available.

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 132–141, 2003.

We extend Viola's approach to include spatial information by proposing a more general similarity measure for the registration problem. Experimental results suggest that our method improves the accuracy and robustness of the registration in the presence of intensity variations. We apply the minimum spanning tree (MST) based entropy estimator [3] and our new registration function to several examples to illustrate the performance of our method. Additionally, we include a comparison of our similarity measure to that used in [4] where joint entropy estimated through an MST algorithm was used as the registration function.

2 Prior Work

Viola et al. [11] and Collignon et al. [5] were the first researchers to use information-theoretic similarity measures for image registration. Mutual information has the advantage of not requiring the definition of landmarks or features such as surfaces. Moreover, it is shown to be well-suited for multi-modal image registration problems.

However, despite some promising results, recent research [7,10,9] has suggested that registration based on mutual information has room for improvement. There can be cases where the mutual information registration function contains local maxima and these can cause misregistration [7]. Different approaches have been proposed to address these issues. These include other information-theory based similarity measures, such as normalized mutual information [10], "higher-order" mutual information [9] and combining gradient information with mutual information [7].

In [7], Puim et al. propose incorporating spatial information by multiplying the registration function with a so-called gradient information term. The main idea is that at registration, locations with large gradient magnitude should be aligned and the orientation of the gradients at those locations should be similar. Experiments suggested that this new measure outperformed regular mutual information based techniques, however the approach is ad-hoc.

In [9], Rueckert et al. suggest employing "second order" mutual information, using the co-occurrence matrices of neighboring voxels' intensities. They showed that second-order mutual-information outperformed first-order mutual information, since it incorporated spatial information into the registration process.

In [4], Hero et al. use the theory of entropic spanning graphs to directly estimate the entropy of a distribution. This has the advantage of by-passing density estimation. Furthermore, for large dimensions this algorithm can be easily implemented, whereas histogram methods cannot. A detailed review of the applications of entropic spanning graphs can be found in [1].

3 Image Registration and Spatial Information

3.1 Model and Main Assumptions

We are given two images: the reference image, $u(x, y)$, and the floating image, $v(x, y)$. We assume the reference image contains the scene represented in the

floating image. Additionally, we assume that each structure (each tissue in medical images) is mapped through different imaging functions for each image to possibly distinct intensities. These mappings may depend on the spatial location. The mathematical relation between the two images is hence modelled as:

$$v(x,y) = f_{x,y}(u(t(x,y))) + W(x,y)\ ,\ \forall (x,y) \in \Omega. \tag{1}$$

In the above equation, $t(x,y)$ represents a geometrical transformation of the coordinates. t may be constrained to a certain class, e.g. affine, rigid-body, spline-based nonlinear, or polynomial transformations, depending on the specific application. Ω is a finite subset of the image plane. $W(x,y)$ is random noise. $f_{x,y}(.)$ is the imaging function; it represents the relation between the pixel intensity values of the two images. The imaging function may depend on the spatial location. In general, we will constrain f by a parameterization as follows: Let $s(x,y)$ be a function from the image plane to $\mathbb{R}^k$, some $k \in \mathbb{Z}^+$, and let $f_{x,y}(.) = g_{s(x,y)}(.)$. Then:

$$v(x,y) = g_{s(x,y)}(u(t(x,y))) + W(x,y),\ \forall (x,y) \in \Omega. \tag{2}$$

3.2 Maximum Likelihood, Conditional Entropy and Registration

In this section we will develop a registration function in a maximum likelihood framework. Let T be a random geometric transformation selected from a given class with respect to a p.d.f. $p_T(t)$. Let U and V be random images, dependent through the transformation T with pixel intensity values satisfying (2). Let $\hat{t}$ be the ML estimate of T given U and V:

$$\hat{t} = \arg\max_t P_{U,V|T}(U = u, V = v | T = t) \tag{3}$$

With the assumption $P_{U|T}(U = u|T = t) = P_U(U = u)$ (3) reduces to

$$\hat{t} = \arg\max_t P_{V|U,T}(V = v | U = u, T = t). \tag{4}$$

Assume the pixel noise $W(x,y)$ is an independent process. It follows from (2):

$$P_{V|U,T}(V = v|U = u, T = t) =$$
$$\prod_{(x,y)\in\Omega} P_{W(x,y)}(W(x,y) = v(x,y) - g_{s(x,y)}(u(t(x,y)))).$$

Hence, maximizing the log-likelihood function of t, is equivalent to maximizing:

$$m(t) = \sum_{(x,y)\in\Omega} \log P_{W(x,y)}(W(x,y) = v(x,y) - g_{s(x,y)}(u(t(x,y)))). \tag{5}$$

However, since the imaging function $g(.)$ and probability distribution of the pixel noise $P_{W(x,y)}(.)$ are unknown, it is not clear how to calculate or estimate $m(t)$. We will investigate the use of conditional entropy to approximate $m(t)$.

Let $N := |\Omega|$, $N_c := |\Omega \bigcap \{(x,y) : s(x,y) = c\}|$ and $N_{a,b,c} := |\Omega \bigcap \{(x,y) : U(t(x,y)) = a, V(x,y) = b, s(x,y) = c\}|$. Note that N and N_c are deterministic values, while $N_{a,b,c}$, since it depends upon U and V, is a random variable. Let $\Delta := \{(a,b,c) : \exists (x,y) \in \Omega, U(t(x,y)) = a, V(x,y) = b, s(x,y) = c\}$. Assume that the pixel noise is identically distributed on every level-set of $s(x,y)$, that is $W(x,y)$ has an arbitrary distribution that satisfies:

$$P_{W(x_1,y_1)}(W(x_1,y_1) = w_1) = P_{W(x_2,y_2)}(W(x_2,y_2) = w_1) = P_{W_c}(w_1) \tag{6}$$

for $\forall\ (x_1,y_1),(x_2,y_2) \in \Omega$ such that $s((x_1,y_1)) = s((x_2,y_2)) = c$.Then, we can rewrite (5) using this alternative notation:

$$m(t) = \sum_{(a,b,c)\in\Delta} N_{a,b,c} \log P_{W_c}(b - g_c(a)). \tag{7}$$

Next, we will introduce some new random variables and relate $m(t)$ to the conditional entropy of these random variables. Let (X,Y) be a uniformly distributed random vector taking values in Ω. For a given transformation t, let $A := U(t(X,Y))$, $B := V(X,Y)$ and $C := s(X,Y)$.

It is obvious that $P_{B|A,C}(B = b|A = a, C = c) = P_{W_c}(b - g_c(a))$. Hence $m(t)$, can be rewritten as:

$$m(t) = \sum_{(a,b,c)\in\Delta} N_{a,b,c} \log P_{B|A,C}(B = b|A = a, C = c). \tag{8}$$

Theorem 1. *$\forall\ (x_1,y_1)$, $(x_2,y_2) \in \Omega$, and $\forall\ t$, assume:*

$$P_{U(t(x_1,y_1))}(U(t(x_1,y_1)) = a) = P_{U(t(x_2,y_2))}(U(t(x_2,y_2)) = a) = P_a$$

Then

$$E(N_{a,b,c}) = N P_{A,B,C}(A = a, B = b, C = c). \tag{9}$$

Proof.

$$E(N_{a,b,c}) = \sum_{(x,y)\in\Omega : s(x,y)=c} P_{U(t(x,y)),V(x,y)}(U(t(x,y)) = a, V(x,y) = b).$$

Applying the assumption on the distribution of the pixel intensity values of U yields $E(N_{a,b,c}) = \sum_{(x,y)\in\Omega:\ s(x,y)=c} P_{V(x,y)|U(t(x,y))}(V(x,y) = b|U(t(x,y)) = a)P_a$. Then using (2), $E(N_{a,b,c}) = N_c P_{W_c}(b - g_c(a))P_a$. Now, let's look at $P_{A,B,C}(.)$.

$$\begin{aligned} P_{A,B,C}(A = a, B = b, C = c) &= P_{B|A,C}(B = b|A = a, C = c)P_{A|C}(A = a|C = c) \\ &\quad \times P_C(C = c) \\ &= P_{W_c}(b - g_c(a))P_a \frac{N_c}{N} = \frac{E(N_{a,b,c})}{N}. \end{aligned}$$ □

The conditional entropy, $H(B|A,C)$ is defined as:

$$H(B|A,C) = -\sum_{a,b,c} P_{A,B,C}(B=b, A=a, C=c)$$

$$\log \frac{P_{A,B,C}(A=a, B=b, C=c)}{P_{A,C}(A=a, C=c)}. \tag{10}$$

Corollary 1. *For $m(t)$ defined in (5):*

$$E(m(t)) = -NH(B|A,C) \tag{11}$$

Proof. From (8): $E(m(t)) = \sum_{\Delta} E(N_{a,b,c}) \log P_{B|A,C}(B=b|A=a, C=c)$. Using (9) and (10) we get: $E(m(t)) = -NH(B|A,C)$. □

Instead of $m(t)$, that is difficult to estimate, we propose using an estimate for $E(m(t)) = -NH(B|A,C)$ as a registration function. This can be estimated using the MST estimator.

3.3 The Minimum Spanning Tree Estimator

In this section, we will provide a brief background on the MST estimator [3]. One advantage of the MST estimator is that it skips density estimation and directly computes an estimate for the entropy. Thus, the complication of fine-tuning the density estimator parameters is avoided. It has also been shown in [2] that this MST estimator has a faster convergence rate compared to plug-in estimators, especially for non smooth densities. Another advantage is that it provides a consistent estimator for higher dimensional feature vectors. Hence, feature vectors that can incorporate spatial or other types of important information can be more easily employed.

In [4], Hero et al. suggest using the Rényi Entropy as a similarity measure for registration. The α-Rényi entropy of a density is defined as:

$$R_\alpha(Z) = \frac{1}{1-\alpha} \log \int_{\mathbb{R}^d} p_Z^\alpha(z) dz \tag{12}$$

where log is the natural logarithm. The limit of Rényi entropy as α approaches 1 is the well-known Shannon differential entropy. Use of Rényi Entropy is motivated by the consistent MST estimator which uses the power weighted length of the minimum spanning tree over the samples from the density. In [4], Hero et al. use raw pixel intensity values obtained from the images as the vertices of the graph, i.e. samples from the density.

Given a set $\chi_n = \{z_1, z_2, ..., z_n\}$ of n points (vertices) in $\mathbb{R}^d$, a *spanning tree* is a connected acyclic graph which passes through all points in χ_n. In this graph, all n points are connected through $n-1$ edges $\{e_i\}$. *The minimum spanning tree* is the spanning tree that minimizes the total weighted edge length $L_\gamma(\chi_n) = \Sigma_i |e_i|^\gamma$ for a fixed γ.

In [3], Hero et al. show that $\hat{R}(Z)$ defined below, is a strongly consistent estimate of the Rényi entropy. Let L_γ be the weighted edge functional and let

$\chi_n = \{z_1, z_2, ..., z_n\}$ be an i.i.d sample drawn from a distribution on $[0,1]^d$ with density $p_Z(z)$. Then

$$\hat{R}(Z) = \frac{1}{1-\alpha}[\log \frac{L_{d(1-\alpha)}(\chi_n)}{n^\alpha} - \log \beta_{L,d(1-\alpha)}],$$

where $\beta_{L,d(1-\alpha)}$ is a constant independent of $p_Z(z)$.

Hero et al. claim that under "proper" registration the Rényi entropy of the joint distribution of features obtained from the overlapped images should be minimized [4,6]. This is reflected as the smallest weighted length of the MST.

In the previous section, we proposed $E(m(t)) = -NH(B|A,C)$ as the registration function. Note that $-NH(B|A,C) = N(H(A,C) - H(A,B,C))$. Let $s : \Omega \to \mathbb{R}^k$. Let Λ and Ψ be lists with repetition defined as $\{(a,c) : \exists (x,y) \in \Omega, U(t(x,y)) = a, s(x,y) = c\}$ and $\{(a,b,c) : \exists (x,y) \in \Omega, U(t(x,y)) = a, V(x,y) = b, s(x,y) = c\}$, respectively. For $\alpha \approx 1$, we suggest the following as an estimate for (11):

$$N[\frac{1}{1-\alpha} \log \frac{L_{(1+k)(1-\alpha)}(\Lambda)}{N^\alpha} - \log \beta_{L,(1+k)(1-\alpha)} - \frac{1}{1-\alpha} \log \frac{L_{(2+k)(1-\alpha)}(\Psi)}{N^\alpha} + \log \beta_{L,(2+k)(1-\alpha)}].$$

Since $\beta_{L,(1+k)(1-\alpha)}$ and $\beta_{L,(2+k)(1-\alpha)}$ don't depend on the transformation t, the registration problem reduces to maximizing:

$$N[\log \frac{L_{(1+k)(1-\alpha)}(\Lambda)}{L_{(2+k)(1-\alpha)}(\Psi)}]. \tag{13}$$

3.4 Spatial Information

Our model incorporates spatial information through the spatial function s. In general, s may not be known. However, depending on the application, one may have good prior knowledge of the function. Note that, the validity of some assumptions such as (6),(2), depends on s. Thus, the more prior knowledge one has on this function, the more realistic is the model.

In an extreme case, one can set $s(x,y) = (x,y)$. Note that, in this case the assumptions on the noise (6) and imaging function (2) are relatively weak. However, intuition and experimental results suggest that the coarser the partitioning defined by s, the more accurate the resulting estimator. In another extreme, one can set $s(x,y) =$ constant. In this case, the assumptions on the noise (6) and imaging function (2) may be too strong.

4 Implementation

The main part of our algorithm involves solving the MST problem on the given set of feature vectors obtained from the images, which are defined to be vectors containing the intensity values and spatial function values, $s(x,y)$ of each

(a) (b)

Fig. 1. a) Functional MR Image b) Floating Image (artificial).

pixel. The Kruskal algorithm is widely believed to be the fastest general purpose algorithm to solve this problem for sparse graphs. It has a time complexity of $O(V^2 \log V)$, V being the number of vertices. This algorithm can be accelerated by applying some modifications suggested by Neemuchwala et al. in [6].

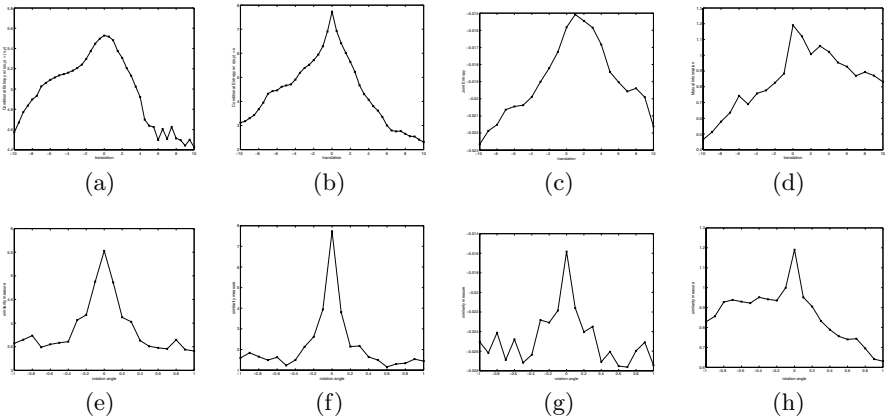
This study concentrates on the accuracy and robustness of the proposed similarity measure in the presence of intensity variances, such as the ones caused by RF inhomogeneity in MR imaging. Thus, the optimization scheme to determine the registration parameters and the transformation space was not included in the study. An exhaustive search was carried out in the registration parameter space to determine the values that maximize the proposed similarity measure. Cubic spline interpolation was used in all experiments.

5 Results

In this section, we demonstrate the performance of our proposed measure in a variety of cases. For comparison, we also illustrate some results obtained by applying the joint entropy measure proposed in [4] and mutual information [11]. In Section 5.1, we demonstrate the behavior of the aforementioned registration functions, in the case of low-resolution multi-modal registration. In Section 5.2 we evaluate the accuracy of the proposed measures by comparison of the registration results against the true values.

5.1 Registration Functions

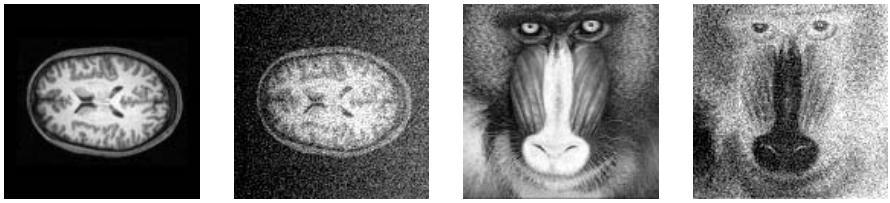
The experiment involves a functional MR image (reference image). The floating image was artificially created by negating the reference image and adding a bias field in the x-direction (see Fig. 1.b). Thus, no transformation is required to align these images. The images have been subsampled by a factor of 8 in each direction. Since there is no random noise, the experimental results characterize the performance of the similarity measures on this pair of images. Figure 2 shows plots of the calculated registration functions. The results suggest that our proposed registration functions are relatively smooth, even with a small number of samples. However, the conditional entropy with restricted spatial information



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 2. Registration Functions. From left to right: Cond. Entropy with $s(x, y) = (x, y)$, Cond. Entropy with $s(x, y) = x$, negative Joint Entropy, Mutual Information. Top row: translation along x-axis, bottom row: rotation around image center.



(a) (b) (c) (d)

Fig. 3. First Experiment: (a) Reference image (structural MR), (b) Floating image. Second Experiment: (c) Reference image, (d) Floating image.

(Fig. 2.b) has a sharper peak compared to the one with full spatial information (Fig. 2.a). This suggests that conditional entropy with spatial information matched to the spatial variation of the imaging function is less susceptible to noise. The joint entropy measure (Fig. 2.c) achieves its optimum at 1 for translation, which is incorrect. It is known that the mutual information registration function is less smooth with a small number of samples, for instance in multi-resolution methods. In the rotation case, conditional entropy with ideal spatial information has the sharpest peak, whereas the joint entropy measure contains many local optima.

5.2 Accuracy

These experiments use the images in Fig. 3. In the first experiment, the floating image was created by applying a bias field in the x-direction and adding i.i.d Gaussian pixel noise to the reference image. This simulates a mono-modal registration experiment in the presence of intensity variances. In the second exper-

Table 1. Estimated translations for the MR (experiment 1) and "baboon" (experiment 2) image pairs with different SNR and bias field magnitudes. The methods are: Conditional entropy (C.E.) with full spatial information (s(x,y)=(x,y)) and ideal spatial information (s(x,y)=x), joint entropy (J.E.) with no spatial information. SNR is defined as the ratio of average pixel intensity to standard deviation of the noise.

		First Exp.				Second Exp.		
Bias	SNR_1	C.E.(s=(x,y))	C.E.(s=x)	J.E.	SNR_2	C.E.(s=(x,y))	C.E.(s=x)	J.E.
0	4.02	0.32	0.50	0	9.45	0	0	0
0.5	4.0	0.39	0.50	3.35	9.45	0	0	1.16
1.0	4.0	1.80	0.50	6.35	9.45	0.22	0.32	0.74
1.5	4.0	2.41	1.32	5.37	9.45	0.45	0.67	0.84
0	2.0	0.39	0.50	0	4.73	0.22	0	0
0.5	2.0	0.63	0.50	2.91	4.73	0	0	1.16
1.0	2.0	2.25	0.50	6.63	4.73	0.55	0.39	1.45
1.5	2.0	1.00	0.50	4.12	4.73	1.50	0.71	1.63
0	1.34	0.50	0.45	0.22	3.15	0.22	0.55	0.32
0.5	1.34	1.00	1.00	2.02	3.15	0.32	0.50	0.84
1.0	1.34	3.27	1.47	6.90	3.15	1.47	0.71	1.02
1.5	1.34	3.81	2.21	6.45	3.15	1.41	0.32	1.83

iment, the reference image was negated before going through the same process to create the floating image. This simulates multi-modal registration. Table 1 presents the results for a variety of noise and bias field magnitudes. The table entries are r.m.s. values of the estimated translations over 5 trials. The true translation is 0.

These results indicate that our proposed method (for both cases: with ideal and non-ideal spatial information) produces a more accurate registration in the presence of intensity variations compared to the joint entropy measure. The experiments also suggest that conditional entropy with spatial information restricted to the prior knowledge of intensity variations is a robust registration function in the presence of high magnitude noise.

6 Future Work

Several issues of the proposed method need further investigation. The robustness of the conditional entropy with restricted spatial information should be researched more extensively. It is known that in MR imaging one may have a strong prior knowledge of the RF inhomogeneity. Our experiments indicate that this knowledge can be used to improve registration. Another important topic is the integration of an optimization scheme to the MST based algorithm. The smoothness of the proposed registration function (even with very low-resolution images) is encouraging. This suggests hill-climbing methods can be applied for the optimization. This is currently under investigation.

Acknowledgements

The MR images were provided by the Center for the Study of Brain, Mind and Behavior at Princeton University. The fast MST code was obtained from Huzefa Neemuchwala and Alfred Hero at University of Michigan Ann Arbor.

References

1. Hero A.O., Ma B., Michel O., Gorman J.D.: Applications of entropic spanning graphs. IEEE Signal Proc. Mag., vol **19**, no **5**, (2002) 85–95
2. Hero A.O., Ma B., Michel O., Gorman J.D.: Alpha divergence for classification, indexing and retrieval. Technical Report CSPL-328 Communications and Signal Processing Laboratory, The University of Michigan, (2001)
3. Hero A.O., Michel O.: Estimation of Rényi Information Divergence via Pruned Minimal Spanning Trees. IEEE Workshop on Higher Order Statistics, Caesaria Israel, (1999)
4. Ma B., Hero A.O., Gorman J., Michel O.: Image registration with minimal spanning tree algorithm. 2000 IEEE International Conf. on Image Processing, Vancouver, (2000)
5. Maes F., Colligon A., Vandermeulen D., Marchal G., Suetens P.: Multimodality image registration by maximization of mutual information. IEEE Trans. Med. Imag., vol. **16**, no. **2**, (1997) 187–198
6. Neemuchwala H.F., Hero A.O., and Carson P.L.: Image registration using alpha-entropy measures and entropic graphs. European Journal on Signal Proc. (2002)
7. Pluim J., Maintz A., Viergever M.: Image registration by maximization of combined mutual information and gradient information. IEEE Trans. Med. Imag., vol. **19**, no.**8**, (2000) 809–814
8. Richard E.B.: Principles and practice of information theory. Reading, MA, Addison-Wesley, (1987)
9. Rueckert D., Clarkson M.J., Hill D.L.G., Hawkes D.J.: Non-rigid registration using higher-order mutual information. Medical Imaging: Image Processing, K.M Hanson, Ed. Bellinghamn, WA, SPIE Press, (2000)
10. Studholme C., Hill D.L.G., Hawkes D.J.: An overlap invariant entropy measure of 3D medical image alignment. Pattern Recognition, vol. **32**, no. **1**, (1999) 71–86
11. Viola P.A., Wells III W.M., Atsumi H., Nakajima S., Kikinis R.: Multi-modal Volume Registration by Maximization of Mutual Information. Medical Image Analysis. vol. **1**, no. **1** (1996) 5–51
12. Viola P.A.: Alignment by maximization of mutual information. Ph.D. Thesis, MIT, Cambridge, MA (1995)

Computerized Atlas-Guided Positioning of Deep Brain Stimulators: A Feasibility Study

Benoit M. Dawant[1], Rui Li[1], Ebru Cetinkaya[1], C. Kao[2], J. Michael Fitzpatrick[1], and Peter E. Konrad[2]

[1] Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, Tennessee
[2] Department of Neurological Surgery, Vanderbilt University, Nashville, Tennessee
Benoit.Dawant@vanderbilt.edu

Abstract. Optimal placement of a deep brain stimulator (DBS) is an iterative procedure. A target is chosen preoperatively based on anatomical landmarks identified on MR images. This point is used as an initial position that is refined intraoperatively using both micro-electrode recordings and macrostimulation. Because the length of the procedure increases with the time it takes to adjust the DBS to its final position, a good initial position is critical. In this work we explore the possibility of using an atlas and non-rigid registration algorithms to select the initial position automatically. We compare the initial DBS position obtained with this approach and the initial position selected by a neurosurgeon with the final position for eight STN (subthalamic nucleus) cases. Our results show that the automatic method leads to initial positions that are closer to the final positions than the initial positions selected manually.

1 Introduction

Since its first FDA approval in 1998 deep-brain stimulation (DBS) has gained significant popularity in the treatment of movement disorders [1,2]. The therapy has significant application in the treatment of tremor, rigidity, and drug induced side effects in patients with Parkinson's disease and essential tremor. The use of a 4-contact electrode, as shown in Figure 1 (Medtronic #3387or #3389 *quadripolar lead*®; Medtronic, Inc., Minneapolis, MN) placed within targets ranging from 4-12 mm in diameter requires stereotactic neurosurgical methodology. Ideally, the optimal target for therapy should be located within the stimulation range of 1 or 2 contacts, each contact measuring 1.5mm separated by either 1.5mm (lead #3387) or 0.5mm (lead #3389). Effective stimulation results when the contacts surround the target [3, 4]. At our institution, we prefer that 2 contacts lie above and 2 contacts lie below the target. If the contacts are located as little as 2mm away from the desired target, ineffective stimulation results due to several reasons: a) failure to capture control of the group of neurons, b) stimulation of non-desirable areas resulting in unpleasant stimulation, or c) necessity for higher stimulus intensities to produce the desired effect resulting in reduced battery life of the implant. For these reasons, targeting the specific neurons of interest for this therapy requires millimetric precision and allowance for variability among patients. Hence, the process of implantation of a DBS electrode follows from a

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 142 –150, 2003.

step-wise progression of a) initial estimation of target localization based on imaged anatomical landmarks, b) intraoperative micro-anatomical mapping of key features associated with the intended target of interest, c) adjustment of the final target of implantation by appropriate shifts in three-dimensional space, and d) implantation of the quadripolar electrode with contacts located surrounding the final desired target. In current clinical practice at our institution, the initial target localization is done manually on MR images based on AC-PC coordinates. In this work we present preliminary data that indicate that the initial target localization could be achieved automatically using non-rigid registration techniques. The data presented herein also indicate that initial target localization predicted by the automatic method is closer or as close to the optimal localization than the initial target localization chosen manually.

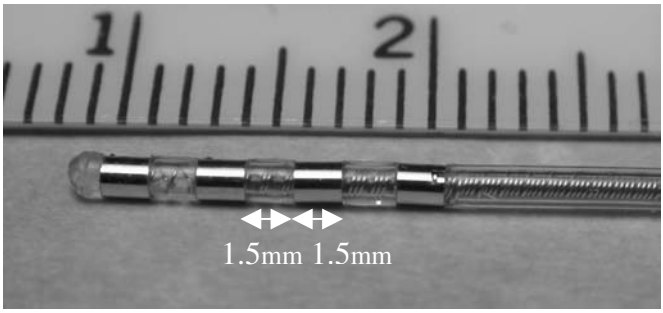


Fig. 1. Medtronic #3387 quadripolar lead® (Medtronic, Minneapolis, MN). Each silver band is one electrode. The numbers on the ruler indicate centimeters.

2 Material and Method

2.1 Patientsa and Preoperative Target Selection

All patients undergoing consideration for DBS implantation of the STN are first evaluated by a movement disorders neurologist and optimized on medications. If patients reach advanced parkinsonian symptoms (rigidity, bradykinesia, tremor, dyskinesias) despite optimal medical therapy, they are considered for surgical therapy by a multi-disciplinary group involving neurology, neurosurgery, neurophysiology, neuropsychiatry specialists. Target selection is decided upon by the team if no contraindications exist. A majority of patients with the above symptoms are recommended for STN targeting of DBS therapy. Target identification is performed by the functional neurosurgeon (PEK) and is based on an identification of the AC-PC location and arriving at 4mm posterior, 12mm lateral, and 4mm inferior to the midcommissural point for STN. Changes in the intended target are modified based on width of the third ventricle and other anatomical asymmetries noted on the MRI scan, but these adjustments usually only consist of less than 1mm deviations from the initial intended target location.

2.2 Guidance System and Surgical Procedure

Traditional methodology for carrying out this stepwise target localization and implantation procedure has been based on an externally fixed, rigid fixture, called a "stereotactic frame" that encompasses the patient's head and upon which the micromanipulating equipment can be mounted and maneuvered with sub-millimetric precision. These various stereotactic frames have been optimized to obtain accurate images used to create the initial target trajectory and plan and then to reduce erroneous movement associated with passage of the test electrodes and the final implant [5]. These frames typically require mounting the day of surgery, subsequent imaging with either CT and/or MRI axial slices, and target planning prior to starting the actual procedure of intraoperative mapping and ultimate placement of the electrode implant into the final target.

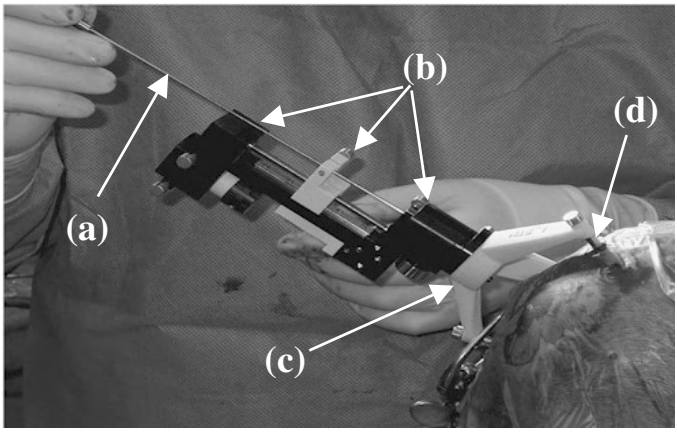


Fig. 2. Surgeon inserting probe (a) into micropositioning drive (b). The drive is attached to the Starfix platform (c). Each leg of the platform is mounted on one marker post, which is outfitted with an adaptor for this purpose (d).

Recently, a market-cleared and CE-compliant miniature stereotactic positioner called a *microTargeting Platform* became clinically available (*microTargeting Drive System for Stereotactic Positioning, incorporating STarFix guidance*, FHC Inc., Bowdoinham, Me.) This device, which we will call a platform, allows for more versatility with elective stereotactic procedures, such as DBS implantation. The platform is currently manufactured as a customized tripod that can be mounted on bone-based fiducial markers. Each platform is uniquely manufactured based on a stereotactically planned trajectory using software designed to mathematically relate the location of such bone markers with respect to brain structures [6]. The bone-based fiducial markers are of a two-piece design in which a fluid-filled cylinder that is visible on both CT and MR is detachably attached to a post that is implanted into the outer table of the skull. These images can then be used in the stereotactic software to designate a trajectory in relation to the bone-based marker posts. The plan is sent to the manufacturer who then translates the stereotactic plan into a customized platform for a given trajec-

tory through a rapid prototyping facility. The resultant platform is shipped to the hospital within 1 week and is used for mounting the same types of micromanipulators that are used on traditional stereotactic frames. The remaining portion of the procedure is the same with respect to intraoperative localization of the final target of implantation with the patient awake.

For each patient, the following data acquistion protocol and preoperative procedure is followed. First, under anesthesia the posts are implanted, Acustar™ (Z-Kat, Inc., Hollywood, FL) fiducial markers[1] are attached to the posts. The use of this marker and post in open craniotomies has been reported on earlier [6]. CT and MR volumes are acquired with the patient anesthetized and head taped to the table to minimize motion. (CT images acquired at kvp = 120V, exposure = 350mas, 512x512 pixels ranging in size from 0.49 to 0.62 mm, slice thickness = 2 mm for one patient, 1.3 mm for 2 patients, 1mm for all others; MR images are 3D SPGR volumes, TR: 12.2, TE: 2.4, voxel dimensions 0.85X0.85X1.3mm^3 except for subject 7 for which the voxel dimensions are 1X1X1.3mm^3). After imaging, the markers are removed. With the help of MR-CT registration software (VoXim^R, FHC, Inc.), the surgeon selects the initial target points based on AC-PC coordinates and associated entry points on the surface of the skull. In addition, the centroids of the markers and the directions of their posts are determined. These data are sent electronically to a fabrication plant where a customized platform is manufactured to fit the posts and provide an opening positioned over the entry point and oriented toward the target.

Surgery begins with the drilling of a burr hole (14 mm). An adaptor is attached to each post, the platforms are attached to the adaptors, and a micropositioning drive (microTargeting^R drive system, FHC Inc., Bowdoinham, ME) is attached to each platform (Figure 2). Micro-electrode recording leads are advanced into the patient to the respective initial target positions through the central tube guide of the platform. Resting firing frequencies are noted and the target positions are revised. The revision involves three-dimensional adjustment. In addition to changes in depth, it is possible to re-insert a probe along parallel tracks distributed within a 10 mm circle around the initial track. The microelectrodes are removed and a unipolar macrostimulation lead is inserted to the revised positions as determined by the micro-electrode recordings. With the patient awake, response to stimulation is monitored as the positions of the probes are further adjusted. When the final positions are selected, the DBS leads are inserted, buried beneath the scalp, and the platform is removed. During the entire procedure coordinates are read on the mircodrive. These physical coordinates can be transformed into preoperative CT coordinates using the software used for preoperative planing.

2.3 Automatic Atlas Creation

The atlas is a common frame of reference in which the position of each individual DBS can be recorded. This requires the spatial normalization of each individual brains. To do so we have used two algorithms developed at our institution. The first one is an independent implementation of the “demons” algorithm proposed by Thirion [8]. The second one is a new algorithm we have developed recently [9] that we call

[1] JMF is a consultant for, and has an interest in, Z-Kat, Inc., Hollywood, FL, which markets this marker.

the Adaptive Basis Algorithm (ABA). In this technique, inspired by the work of Rueckert *et al.* [10] and Meyer *et al* [11], the deformation that registers one image onto the other is modeled with a linear combination of radial basis functions with finite support. The similarity measure used to drive the registration process is the mutual information between the images. In this algorithm, several improvements over existing mutual information-based non-rigid registration algorithm are implemented. These include working on an irregular grid, adapting the compliance of the transformation locally, decoupling a very large optimization problem into several smaller ones, and deriving schemes to guarantee the topological correctness of the transformations. This algorithm computes the final transformation iteratively across scales and resolutions (in this context, resolution means the spatial resolution of the image while the scale is related to the transformation itself). A standard image pyramid is created to apply the algorithm at different resolutions. At each resolution, the scale of the transformation is adapted by modifying the region of support and the number of basis functions (the scale of the transformation is indeed related to the bases' region of support; a large region of support leads to a transformation at a large scale). Typically the algorithm is initialized on a low resolution image with few basis functions having large support. As the algorithm progresses to finer resolutions and smaller scales, the region of support of the basis functions is reduced. Following this approach, the final deformation field is computed as

$$\mathbf{v}(\mathbf{x}) = \mathbf{v}_1(\mathbf{x}) + \ldots + \mathbf{v}_M(\mathbf{x}).$$

with M the total number of levels with one level referring to a particular combination of scale and resolution. In this study we have chosen empirically one of the MR volumes as the atlas. All the other MR volumes are then registered automatically to the atlas using both algorithms. Once the transformation between one image volume and the atlas is computed, the spatial coordinates of the DBS in this volume can be transformed into atlas coordinates. The optimun DBS position in the atlas is computed as the centroid of all the DBS positions after their projection onto the atlas. But because the intraoperative coordinates are given in terms of preoperative CT coordinates and because the non-rigid registration algorithms need to be applied on MR images, an additional step is required. Corresponding MR and CT image volumes are registered using a rigid body transformation also computed using mutual information as proposed by Maes et al. [12].

2.4 Prediction of the Optimum Target Position with the Atlas

Predicting the DBS position for each patient is the inverse of the operation described above. It consists in projecting the optimum DBS position from the atlas to each individual image volume. This does not require another registration step because we compute the transformation from the patient to the atlas and from the atlas to the patient simultaneously. Our algorithms impose constraints on these transformations to keep them almost inverse of each other to produce bijective transformations.

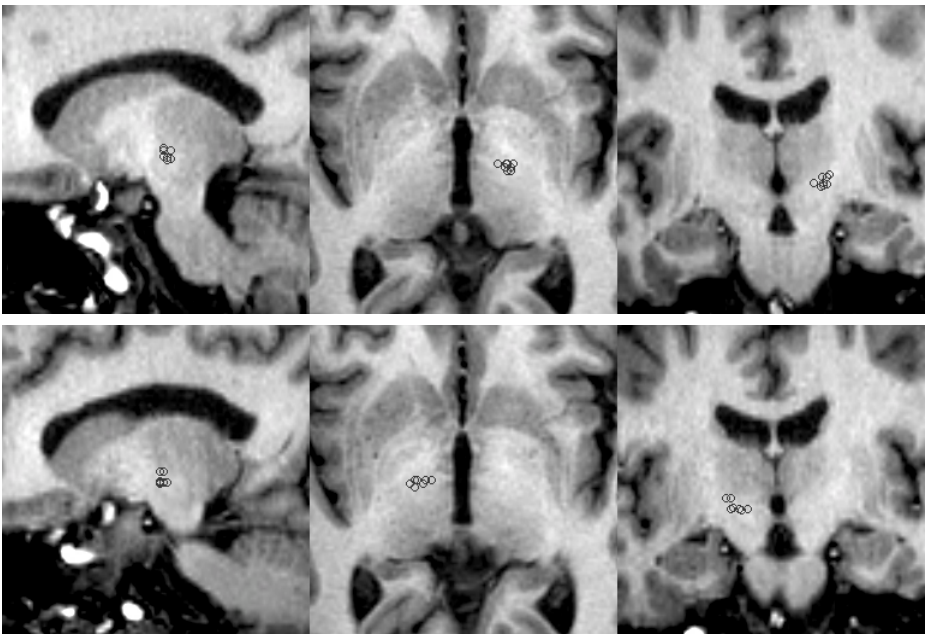


Fig. 3. Final location of the DBSs in the atlas projected onto the sagital, transverse, and coronal planes passing trough their centroid; top left side; bottom right side.

3 Results

3.1 Visual Evaluation of the Registration Results

Figure 3 shows the location of each of the DBSs in the atlas created with the ABA algorithm (results obtained with the demons algorithm are qualitatively similar). To generate these figures the position of these points has been projected on the sagital, transverse, and coronal planes passing through their centroid. The top panels show the left side, the bottom panels the right side.

3.2 Projection of the Final DBS Positions onto the Atlas

Tables 1 and 2 list the coordinates of the final DBS position transformed into atlas coordinates for eight bilateral STN patients using both algorithms. The centroid for the left and right sets of points has been computed. The Euclidean distance between each point and its corresponding centroid is reported in the Dc column. The distance between the left centroids computed with the ABA algorithm and the demons algorithm is 1.22mm. The distance between the right centroid computed with the ABA algorithm and the demons algorithm is 1.16mm. These results show that the final position of the DBSs transformed into atlas coordinates result in tight clusters. It is also worth noting that even though these two algorithms are based on very different

similarity measures (one minimizes the intensity difference between images on a voxel-by-voxel basis, the other maximized the Mutual Information between the volumes) they lead to essentially identical results.

Table 1. Position of the DBS position transformed into atlas coordinates using the ABA algorithm. Dc refers to distance from centroid

Adaptive Basis Algorithn								
	Left (all values are in mm)				*Right (all values are in mm)*			
	X	Y	Z	Dc	X	Y	Z	Dc
S1	124.28	106.39	53.98	***2.58***	95.47	106.20	53.20	*3.52*
S2	121.88	106.30	53.44	***1.44***				
S3	122.47	107.80	51.18	***1.43***	96.86	107.02	50.24	*2.12*
S4	119.79	106.27	51.78	***2.74***	101.38	104.91	50.23	*3.40*
S5	123.42	107.41	51.40	***1.42***	99.10	105.99	50.19	*1.35*
S6	123.42	107.99	53.21	***1.82***	96.68	104.69	52.96	*2.57*
S7	122.48	105.80	51.90	***1.07***	97.54	104.98	50.51	*0.92*
S8	121.70	106.60	50.91	***1.52***	99.85	104.69	49.63	*2.34*
Mean	*122.43*	*106.82*	*52.22*	*1.75*	*98.12*	*105.49*	*51.00*	*2.32*
STD	*1.37*	*0.81*	*1.16*	*0.60*	*2.07*	*0.91*	*1.45*	*0.97*

Table 2. Position of the DBS position transformed into atlas coordinates using the demons algorithm. Dc refers to distance from centroid.

Demons Algorithm								
	Left (all values are in mm)				*Right (all values are in mm)*			
	X	Y	Z	Dc	X	Y	Z	Dc
S1	124.28	106.39	53.98	*2.29*	95.47	106.20	53.20	*2.90*
S2	121.60	104.24	54.43	*1.94*				
S3	122.03	106.52	53.04	*0.88*	95.92	105.95	51.87	*2.14*
S4	120.31	106.73	51.97	*2.54*	99.78	105.72	51.23	*2.21*
S5	123.11	106.00	52.77	*1.08*	98.85	104.68	51.94	*1.08*
S6	122.42	105.63	54.60	*1.32*	97.01	103.79	53.97	*2.44*
S7	121.86	104.44	53.44	*1.32*	97.89	104.19	52.34	*0.89*
S8	122.07	105.63	52.16	*1.15*	99.92	104.58	49.61	*3.22*
Mean	*122.21*	*105.70*	*53.30*	*1.56*	*97.83*	*105.01*	*52.02*	*2.12*
STD	*1.15*	*0.93*	*0.99*	*0.61*	*1.78*	*0.94*	*1.40*	*0.87*

3.3 Comparison between Manually Selected and Atlas-Suggested Initial Target Positions

Table 3 presents the Euclidean distance computed between the final DBS position selected intraoperatively and (a) the initial position chosen manually and preoperatively by the neurosurgeon, (b) the initial position suggested by projecting the DBS position from the atlas onto each subject using the ABA algorithm and (c) the same as in (b) but with the demons algorithm.

Table 3. Distance between the initial position selected manually and automatically and the final position selected intraoperatively

Distances between original and final DBS position (in mm)							
	Left				***Right***		
Subject	Manual	ABA	Demons		Manual	ABA	Demons
S1	5.95	2.58	2.40		6.94	3.52	2.94
S2	5.72	1.78	3.39				
S3	2.53	2.41	1.50		4.49	2.45	2.23
S4	5.30	2.46	2.80		1.99	3.24	2.36
S5	2.31	2.17	1.00		3.64	1.60	1.51
S6	5.95	2.36	2.85		7.31	3.39	3.84
S7	2.00	1.64	2.37		2.01	1.75	1.68
S8	1.71	1.75	0.70		1.67	2.94	3.67
Mean	*3.93*	*2.14*	*2.13*		*4.01*	*2.70*	*2.60*
Std	*1.94*	*0.37*	*0.96*		*2.36*	*0.78*	*0.91*

4 Conclusions and Discussion

The results presented in this paper, albeit based on a small number of subjects, indicate that a fully automatic method for DBS target identification is possible. With both algorithms, the final position of the DBS, when mapped onto the atlas lead to tight clusters with average point-to-centroid distance in the order of 1.5 voxel. With both algorithms, the initial target points also are substantially closer to the final ones than the initial target point chosen manually (the average distance between initial and final position is 45% smaller with the automatic method on the left side and 30% on the right). Despite the small size of our data set the distance between the initial target points and the final target points is significantly smaller ($P<0.01$, one sided paired t-test) than the distance between the initial target points chosen manually and the final target points for both algorithms on the left side. On the right side, the significance drops to ($P<0.07$) and ($P<0.06$) for the ABA and the demons algorithms, respectively.

Atkinson et al. [13] have also explored the idea of using an atlas for movement disorder related surgery. This group correlated the clinical efficacy of stereotactic thalamotomy for tremor with anatomical localization by using postoperative magnetic resonance (MR) imaging and a deformable atlas of subcortical structures. These authors have been able to demonstrate a significant difference in the position of the lesion in their atlas for patients in three clinical outcome groups: excellent, good, and fair. However, they do not provide data in which the position of the lesion predicted by the atlas can be quantitatively compared to either the initial position selected by the neurosurgeon or the final position chosen intraoperatively. Other differences include the fact that their procedure is performed with a stereotactic frame and that they rely on a lesion to eliminate the tremor rather than an implantable stimulator.

A number of issues remain to be investigated. Because the number of patients for which we have the necessary data is limited, we have evaluated our approach on the set used to create our atlas. This may bias the results in our favor. As the number of data sets increase, we will separate the volumes into training and testing set to address

this issue. The best way to develop the atlas also remains an area of investigation. In the current study we have chosen one image volume as our atlas. We have not studied the impact of this choice on the results. We have also used all the image volumes we had at our disposal regardless of clinical outcome. A better approach may be to select only cases for which the clinical outcome is excellent to build the atlas.

If verified, the results presented herein may have a significant impact on the availability of the procedure. It is estimated that in the US alone 10-20,000 patients would benefit from DBS implantation each year. This number of procedures cannot be performed in leading research institution alone in which neurosurgeons have years of experience selecting targets manually. It is hoped that computer-assistance in target identification might make this procedure easier to perform by less experienced surgeons and hence make it available to many patients to whom it would otherwise remain inaccessible.

References

1. Referen G. Deuschl, J. Volkmann, P. Krack, "Deep brain stimulation for movement disorders", *Movement Disorders*, vol.17 (supplement 3) pp, S1-S1, 2002.
2. B. Schrader, W. Hamel, D. Weinert, H. M. Mehdorn, "Documentation of electrode localization.", *Movement Disorders*, vol. 17 (supplement 3), pp S167-S174, 2002.
3. Vitek, J.L., Mechanisms of deep brain stimulation: excitation or inhibition. *Mov Disord*, 2002. 17 Suppl 3: p. S69-72.
4. Lozano, A.M., Deep brain stimulation for Parkinson's disease. 2001. 7(3): p. 199-203.
5. R. L. Galloway and R. J. Maciunas, "Stereotactic neurosurgery", *Crit Rev Biomed Eng* 18(3), pp. 181-205, 1990.
6. J. Franck, P. Konrad, R. Franklin, F. Haer, D. Hawksley. "STarFix: A Novel Approach to Frameless Stereotactic Neurosurgery Utilizing a Miniaturized Customized Pretargeted Cranial Platform Fixture – Technical Description, Unique Features, and Case Reports", *Movement Disorders Society*, 7th Intl. Congress of Parkinsons Disease & Movement Disorder, Miami, FL, November 2002.
7. C. R. Maurer, Jr., J. M. Fitzpatrick, M. Y. Wang, R. L. Galloway, Jr., R. J. Maciunas, and G. S. Allen, "Registration of head volume images using implantable fiducial markers," *IEEE Trans. Med. Imaging,* vol. 16, pp. 447-462, 1997.
8. J.-P. Thirion, "Image matching as a diffusion process: an analogy with Maxwell's demons*". Medical Image Analysis*, vol. 2, no. 3, pp. 243-260, 1998.
9. G. Rhode, A. Aldroubi and B. M. Dawant, "The Adaptive-bases algorithm for intensity-based nonrigid image registration," *IEEE Transactions on Medical Imaging* (in press, 2003)
10. D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Non-rigid Registration Using Free-Form Deformations: Application to Breast MR Images." *IEEE Transactions on Medical Imaging*, 18(8), pp 712-721, 1999.
11. C. R. Meyer et al., "Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate" *Medical Image Analysis*, vol. 3, pp. 195-206, 1997
12. F. Maes, A. Collignon, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Transaction on Medical Imaging* 16(2), pp. 187–198, April 1997.
13. J.D. Atkinson, D.L. Collins, G. Bertrand, T.M. Peters, G.B. Pike, and A.F. Sadikot, "Optimal location of thalamotomy lesions for tremor associated with Parkinson Disease: a probabilistic analysis based on postoperative magnetic resonance imaging and an integrated digital atlas*", J. Neurosurgery*, 96, pp. 854-866, 2002.

Evaluation of Intensity-Based 2D-3D Spine Image Registration Using Clinical Gold-Standard Data

Daniel B. Russakoff[1,2], Torsten Rohlfing[2], Anthony Ho[3], Daniel H. Kim[2], Ramin Shahidi[2], John R. Adler, Jr.[2,3], and Calvin R. Maurer, Jr.[2]

[1] Department of Computer Science
[2] Image Guidance Laboratories, Department of Neurosurgery
[3] Department of Radiation Oncology
Stanford University, Stanford, CA, USA
dbrussak@stanford.edu,calvin.maurer@igl.stanford.edu

Abstract. In this paper, we evaluate the accuracy and robustness of intensity-based 2D-3D registration for six image similarity measures using clinical gold-standard spine image data from four patients. The gold-standard transformations are obtained using four bone-implanted fiducial markers. The three best similarity measures are mutual information, cross correlation, and gradient correlation. The mean target registration errors for these three measures range from 1.3 to 1.5 mm. We believe this is the first reported evaluation using clinical gold-standard data.

1 Introduction

In order to use preoperatively acquired three-dimensional (3D) images for intraoperative therapy guidance, the images must be registered to a patient coordinate system defined in the operating room. Image-to-physical registration is one of the fundamental steps in all image-guided interventions. Surgical navigation systems use the image-to-physical registration transformation to track in real time the changing position of a surgical probe on a display of the preoperative images or to direct a needle to a surgical target visible in the images. Stereotactic radiotherapy and radiosurgery systems use the image-to-physical transformation to direct radiation to a surgical target visible in the images.

A promising method for obtaining the image-to-physical transformation is the registration of a 3D x-ray computed tomography (CT) image to one or more two-dimensional (2D) x-ray projection images (e.g., fluoroscopy images, amorphous silicon detector images). This approach has applications in image-guided spine surgery [1,2] and radiosurgery [3,4]. The 2D-3D registration problem involves taking one or more x-ray projection (2D) images of the patient's anatomy and using these projections to determine the rigid transformation $\mathbf{T}$ (rotation and translation) that aligns the coordinate system of the CT (3D) image with that of the x-ray projection images and the operating room. Figure 1 shows a schematic representation of the 2D-3D registration process. In general, most of the proposed solutions to this problem fit into this framework.

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 151–160, 2003.

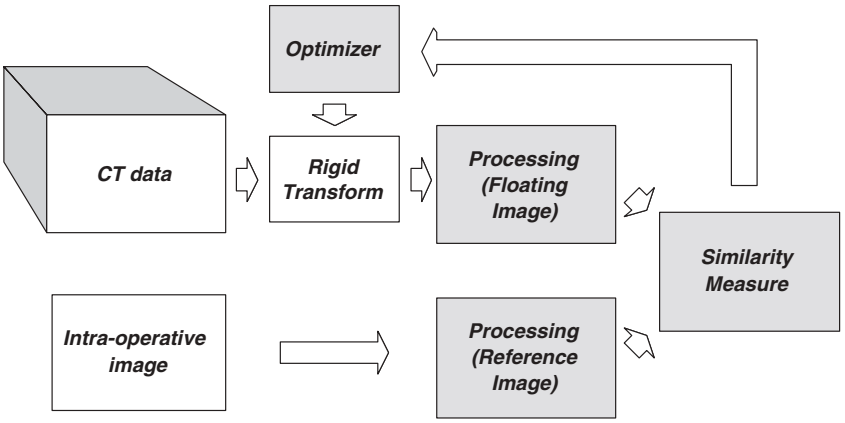
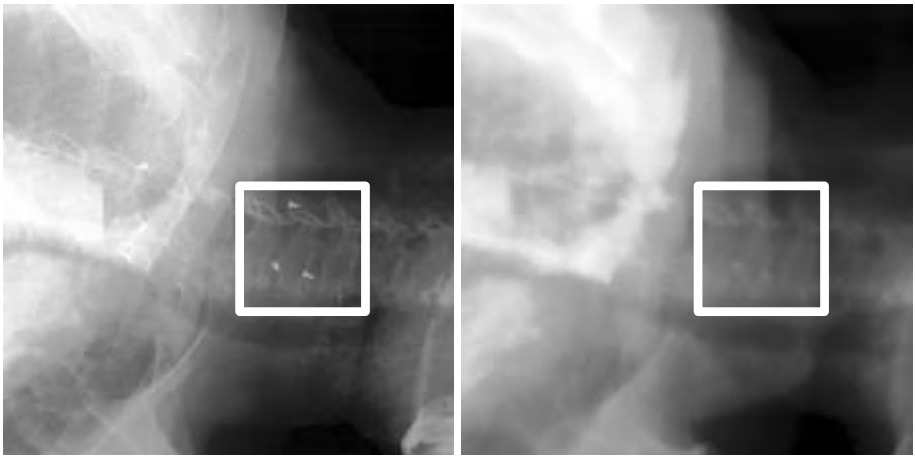


Fig. 1. Schematic overview of the 2D-3D registration process. For intensity-based 2D-3D registration, the reference image is an intra-operative x-ray projection (2D) image. It is used as is with little or no processing. The floating image is a CT (3D) image. It is processed by generating DRRs (synthetic x-ray projection images) for various orientations of the CT image relative to the reference frame of the x-ray imaging system. The optimizer searches for the rigid transformation **T** that produces the DRR that is most similar to the real x-ray projection image. The optimal transformation is used to align the CT coordinate system with that of the operating room.

We are particularly interested in intensity-based 2D-3D image registration [2,5,6,7]. In this case, the reference image is one or more x-ray projection images and the floating image is a CT image. The method involves computing synthetic x-ray images, which are called digitally reconstructed radiographs (DRRs), by casting rays using a known camera geometry through the CT image. The DRR pixel values are simply the summations of the CT values encountered along each projection ray. The pose (position and orientation) of the CT image (given by the transformation **T**) is adjusted iteratively until the DRR produced is most similar to the x-ray projection image. Figure 2 shows a corresponding pair of real and synthetic x-ray projection images.

A variety of similarity measures have been used, including cross correlation, pattern intensity, gradient correlation, gradient difference, entropy, and mutual information [6]. Intensity-based 2D-3D image registration appears to be more accurate than feature-based registration (not including metal fiducial markers) [8]. But for spine images, intensity-based 2D-3D registration methods are very susceptible to local minima in the cost function and thus need initial transformations that are close to the correct transformation in order to converge reliably [6]. Also, intensity-based registration methods are basically untested on gold-standard clinical spine image data.

In this paper, we evaluate the accuracy and robustness of intensity-based 2D-3D registration for the six aforementioned image similarity measures using clinical gold-standard spine image data from four patients. The gold-standard transformations are obtained using four bone-implanted fiducial markers. We believe this is the first reported evaluation using clinical gold-standard data.



(a) (b)

Fig. 2. (a) Example x-ray projection image. A box delineates a region of interest (ROI). Three bone-implanted fiducial markers can be seen in the ROI. (b) A DRR (synthetic projection image) generated from a registered (the correct position and orientation have already been calculated) CT image of the same patient using the camera geometry of the x-ray imaging system used to create the image in (a).

2 Gold-Standard Clinical Spine Image Data

The CyberKnife Stereotactic Radiosurgery System (Accuray, Inc., Sunnyvale, CA) is an image-guided frameless robotic stereotactic radiosurgery system that was developed as a noninvasive means to precisely align treatment beams with targets [9]. Two orthogonal amorphous silicon detector (ASD) x-ray cameras in the treatment room establish a coordinate frame to locate the patient's target site with respect to the therapy beam directions for the robotic manipulator (Fig. 3). A pair of images from the camera system determines the patient's position during treatment. Because the treatment position can differ from the position in the CT planning study, a 2D-3D image registration process is used to find the rigid-body transformation that relates the CT position to the treatment position. This transformation is communicated through a real-time control loop to a robotic manipulator that points a compact 6 MV x-band linear accelerator (LINAC). By taking images throughout the treatment process, shifts in patient position can be detected and the beams can be redirected accordingly.

Patients are currently undergoing treatment of spinal cord lesions with the CyberKnife through an extended FDA treatment protocol for use of the device. Before treatment, each patient is fitted with a simple immobilization device. The cervical spine patients are fitted with a molded Aquaplast face mask (WFR/Aquaplast Corp., Wyckoff, NJ) that stabilizes the head and neck on a radiographically transparent headrest. Thoracic and lumbar spine patients rest in a conformal alpha cradle during CT imaging and treatment. These supports maintain the general orientation of the anatomy and minimize patient motion.

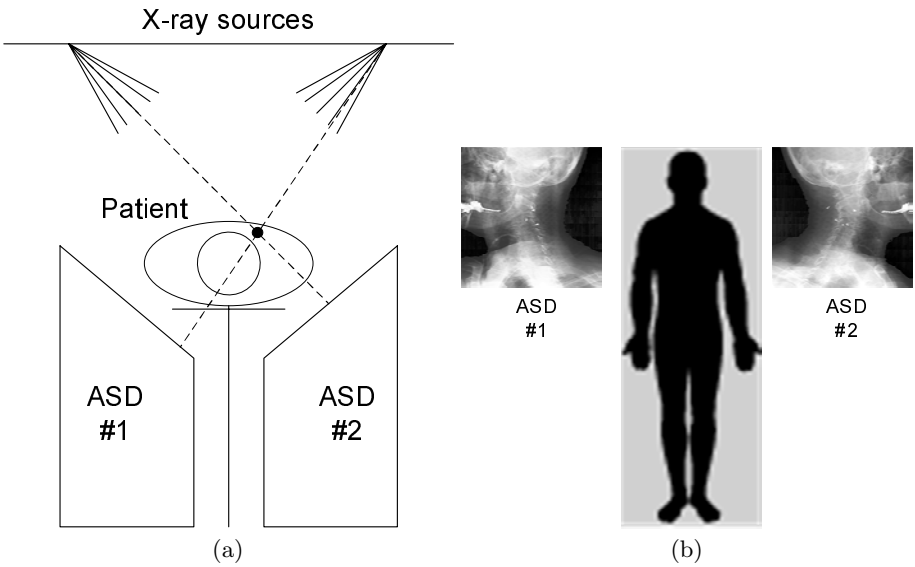


Fig. 3. (a) Schematic illustration of the CyberKnife system looking towards the head of a patient lying on the treatment table. Two x-ray sources are mounted on the ceiling. Two amorphous silicon detector (ASD) x-ray cameras are mounted on the floor. The orthogonal x-ray projection imaging system is calibrated so that the intrinsic and extrinsic cameras parameters of both imaging devices are known. The dashed lines indicate the triangulation of the 3D position of a fiducial marker that appears in both x-ray projection images. (b) A birds-eye view of the system showing a patient lying on the treatment table with two x-ray projection images. In these images, the fiducial markers appear as small, white dots in the cervical vertebrae near the base of the skull.

For each patient, a contrast CT scan of the region of interest is acquired for treatment planning and also for reference in the image-guidance process. For spinal radiosurgery of thoracic and lumbar vertebrae, and most cervical vertebrae, fiducial markers are implanted percutaneously before CT scanning in the posterior bony elements of the vertebral levels adjacent to the lesions to provide radiographic landmarks. Because these implanted fiducials have a fixed relationship with the bone in which they are implanted, any movement in the vertebrae is detected and compensated by the CyberKnife. Implantation of fiducial markers occurs in the operating room under conscious sedation. The fiducial markers are 2×6 mm surgical stainless steel self-retaining tacks. Four or more fiducials are placed in a noncoplanar pattern and spaced approximately 25 mm apart. Three non-collinear fiducials are required to define a rigid-body transformation. Four fiducials provide redundancy in the event that one of them is obscured or otherwise difficult to image. Each fiducial is implanted through stab wounds in the skin and guided with intraoperative fluoroscopy. They are implanted in the lamina or facet of the spine around the lesion of interest. No complications have been reported from this procedure, and all patients have been discharged home the same day.

We obtained archived CyberKnife spinal image data from four patients. Two of these patients have cervical vertebrae lesions (C3 and C5) and two have thoracic vertebrae lesions (T1 and T8). For each patient, we obtained: 1) A pretreatment CT image with slice thickness 1.25 mm and a field of view sufficiently large to image the entire cross section of the body. 2) Approximately 20–30 pairs of orthogonal projection x-ray images obtained at intervals of approximately 60 seconds for the duration of treatment with the two Flashscan 20 flat-panel amorphous silicon x-ray cameras (dpiX, LLC, Palo Alto, CA). The x-ray images have 512×512 pixels with pixel size 0.4 mm and 12-bit intensity values. Only one randomly chosen pair of projection x-ray images is used for the work reported in this paper. 3) The camera calibration model and parameters for the two x-ray cameras. These parameters are obtained by scanning a calibration phantom as part of regular quality assurance testing. 4) Positions (3D) of the four fiducial markers in the CT image. 5) Positions (2D) of the four fiducial markers in the projection x-ray images.

2.1 Assessment of Registration Accuracy and Robustness

A gold-standard reference transformation is determined as follows. Each pair of corresponding 2D projection x-ray fiducial positions is backprojected to reconstruct the 3D fiducial position. The rays do not generally intersect. We take as the 3D coordinate the midpoint of the shortest line segment between the two rays. Then we perform a point-based registration by finding the rigid transformation that aligns the 3D fiducial positions from the CT image with the 3D backprojected fiducial positions from the x-ray images, such that the distance between corresponding points is minimized in the root-mean-square sense. The target registration error (TRE) of a registration transformation being evaluated is computed as the difference between the positions of a target mapped by the evaluated transformation and the gold-standard transformation [10]. The TRE values are computed for each voxel inside a rectangular box bounding the vertebra.

3 2D-3D Image Registration Algorithm

The algorithm searches for the six parameters of the rigid transformation that produces the DRR (synthetic x-ray projection image) that is most similar to the real x-ray projection image. The algorithm performs four main functions corresponding to the four shaded boxes in Fig. 1: processing of the reference image, processing of the floating image, computation of a similarity measure, and optimization.

3.1 Reference Image

We crop the reference image to include a specific region of interest (ROI) (Fig. 2). The ROI includes the anatomy that will be treated. Restricting the registration

to a ROI has several advantages. First, this speeds up the registration process. In particular, DRRs are computationally expensive to create, and their generation is typically a bottleneck in the execution of the registration process. The DRRs are generated only for the ROI. Also, the similarity measure is computed only for the ROI. Second, the registration should be more accurate within the ROI. The smaller the ROI, the less likely that structures within the ROI have moved relative to each other between the time the preoperative CT is acquired and the time the procedure is performed. The definition of the ROI is performed manually and requires minimal effort. We generally specify an ROI that includes a vertebra of interest plus the two adjacent vertebra.

3.2 Floating Image

A DRR (synthetic x-ray projection image) is generated for each transformation considered during the iterative search process. We use a light field rendering method to generate the DRRs. Using light fields allows most of the computation to be performed in a preprocessing step. After this precomputation step, very accurate DRRs can be generated quickly (about 50 ms for a 256×256 DRR). The details can be found in Ref. [11].

3.3 Image Similarity Measures

We perform registrations using the six image similarity measures that are described and used in Penney *et al.* [6]. Specifically, the six similarity measures are: cross correlation, pattern intensity, gradient correlation, gradient difference, difference image entropy, and mutual information.

3.4 Optimization Strategy

We currently use a fairly simple best neighbor search strategy similar to that in Ref. [12]. Basically the search process takes an initial transformation $\mathbf{T}_0$ as input. The twelve closest neighbors in parameter space are computed by varying each transformation parameter by some given step size. There are twice as many neighbors as parameters because the step size is both added and subtracted in order to look in both directions. Each neighbor $\mathbf{T}_i$ is itself a transformation and is evaluated by generating DRRs using $\mathbf{T}_i$ and the geometry of each x-ray camera and computing the similarity between the DRRs and the reference x-ray projection images.

The neighbor with the best value of the cost function is picked, its neighbors examined, and so on until no further improvement in the value of the cost function can be made for the current step size. Then the process is repeated using the current best transformation as the initial transformation and a smaller step size, which is half of the previous step size. This continues until some predetermined resolution is reached. The parameter step sizes are normalized using a scaling factor such that for a given step size, the average motion of all projected voxels in the projection plane is approximately equal for all parameters [13].

Table 1. 2D-3D Spine Image Target Registration Error.

Similarity Measure	TRE (mm)						Unsuccessful Registrations
	Pat. 1	Pat. 2	Pat. 3	Pat. 4	Mean	Max	
Mutual Information	1.5	1.7	0.9	1.2	1.3	1.7	6%
Cross Correlation	1.3	1.8	1.0	1.7	1.5	1.8	6%
Gradient Correlation	2.2	1.0	1.1	1.0	1.3	2.2	15%
Pattern Intensity	1.8	1.2	2.0	1.4	1.6	2.0	38%
Gradient Difference	2.0	0.9	1.2	1.1	1.3	2.0	23%
Diff. Image Entropy	2.1	1.9	2.0	1.7	1.9	2.1	57%

The search is performed in two passes, the first with smoothed versions of the reference images, and the second with the actual reference images. The reference images are smoothed in the first pass using a Gaussian filter with $\sigma = 1.5$ mm. This procedure has the effect of smoothing the cost function in order to help avoid local minima and to produce a good initial transformation for the second step. This is a multi-scale search strategy rather than a multi-resolution search strategy, which we have used in previous 3D-3D image registration work. Because the 2D images have a relatively limited number of pixels, we opt to blur the images and use all of the pixels rather than subsample the images in order to better estimate the joint probability density function and joint entropy. The initial step size for the first pass corresponds to an average motion of projected voxels of 5 mm. This is successively decreased to a final step size of 0.5 mm. The initial and final step sizes for the second pass are 2 mm and 0.1 mm, respectively.

During the iterative search process, we use a CT image origin that is centered in the region of interest. The selection of this origin is performed manually and requires minimal effort. For spine image registration, we generally specify an origin that is a point in the center of the vertebral body of interest.

4 Results

Initial transformations were generated by perturbing the gold-standard reference transformation by adding randomly generated rotations and translations. The initial transformations were characterized by computing the TRE for the transformation and grouped into six initial TRE intervals: 0–2, 2–4, 4–6, 6–8, 8–10, and 10–12 mm. For each patient and each similarity measure, 480 registrations were performed, 80 in each of the six misregistration intervals. The TRE value was computed for each registration transformation as the difference between the positions of a target mapped by the evaluated transformation and the gold-standard transformation. The TRE values were computed for each voxel inside a rectangular box bounding the vertebra and then averaged. The registrations were characterized as either "successful" if the TRE < 2.5 mm or "unsuccessful" if the TRE ≥ 2.5 mm. The results are listed in Table 1. The TRE values that are listed are the mean TRE for all successful registrations.

The pattern intensity similarity measure has two user-defined parameters, r and σ. We evaluated pattern intensity using several values of these parameters.

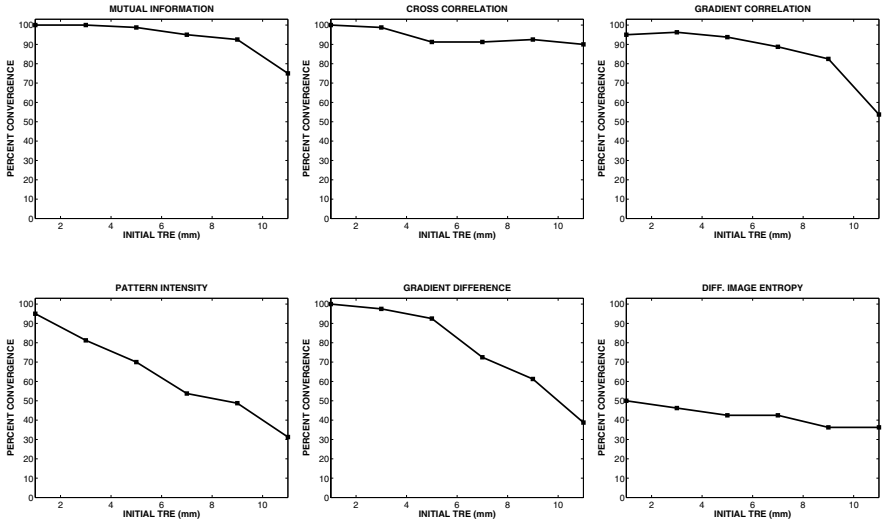


Fig. 4. Percentage of successful registrations for initial transformations with different initial TRE values for six image similarity measures. Each data point represents a 2 mm range of initial TRE values centered at the x-coordinate.

The results we list in Table 1 are the best results we obtained, and correspond to $r = 4$ mm (10 pixels) and $\sigma = 10$.

Based on the results listed in Table 1, mutual information, cross correlation, and gradient correlation outperformed the remaining three similarity measures. The mean TRE ranges from 1.3–1.5 mm for these three similarity measures, and the frequency of unsuccessful registrations ranges from 6–15%. With only four patients, it is not possible to really distinguish the performance of these three measures from each other. Gradient difference has similar mean and max TRE values, but 23% of the registrations failed. Two of the similarity measures, pattern intensity and difference image entropy, performed substantially worse than the three best measures. The frequency of unsuccessful registrations for these two measures is 38% and 57%, respectively.

Figure 4 shows how the percentage of successful registrations depends on the accuracy of the initial transformation. The three best similarity measures based on the results listed in Table 1 are also the three best based on the results shown in this figure. All three measures have a high probability of successful registration as long as the initial transformation has a TRE < 8 mm, with decreasing robustness as the initial transformation gets further from the correct transformation. Pattern intensity and gradient difference have a high probability of successful registration only if the initial transformation has a TRE < 4 mm. Difference image entropy has a poor probability of successful registration regardless of the accuracy of the initial transformation.

5 Discussion

The results in this paper are substantially different than those reported by Penney *et al.* [6]. For example, we found that mutual information is one of the best similarity measures, whereas they found that it is one of the worst. One possible explanation is that we use higher resolution images in our calculations. As a result, our ROIs have more pixels, which provides more accurate estimates of probability density functions and entropies. We also found that pattern intensity produced relatively poor results with our data, whereas Penney *et al.* found it to be an accurate and robust measure. We note that the three worst similarity measures in our study, pattern intensity, gradient difference, and difference image entropy, are all computed using a difference image. One of the most important differences between the Penney *et al.* study and this work is that they used a phantom and we used clinical data. It is possible that different similarity measures produce different results with different kinds of data.

It is difficult to make strong conclusions with limited data. The results in this paper are based on image data from only four patients, two with lesions in cervical vertebrae and two with lesions in thoracic vertebrae. We have access to clinical gold-standard image data from at least twenty additional patients and plan to analyze this data in the future. These patients have spinal lesions at vertebral locations distributed throughout the spinal column, including the lumbar region.

Acknowledgments

DBR was supported by the Interdisciplinary Initiatives Program, which is part of the Bio-X Program at Stanford University, under the grant "Image-Guided Radiosurgery for the Spine and Lungs." TR was supported by the National Science Foundation under Grant No. EIA-0104114. The authors thank John Dooley, Gopinath Kuduvalli, and Matt Core (Accuray, Inc., Sunnyvale, CA) for their technical assistance.

References

1. S. Lavallee, J. Troccaz, P. Sautot, B. Mazier, P. Cinquin, P. Merloz, and J.-P. Chirossel, "Computer-assisted spinal surgery using anatomy-based registration," in *Computer-Integrated Surgery: Technology and Clinical Applications*, R. H. Taylor, S. Lavallee, G. Burdea, and R. Mösges, Eds. Cambridge, MA: MIT Press, 1996, pp. 425–449.
2. J. Weese, G. P. Penney, T. M. Buzug, D. L. G. Hill, and D. J. Hawkes, "Voxel-based 2-D/3-D registration of fluoroscopy images and CT scans for image-guided surgery," *IEEE Trans. Inform. Technol. Biomedicine*, vol. 1, pp. 284–293, 1997.
3. M. J. Murphy, J. R. Adler, Jr., M. Bodduluri, J. Dooley, K. Forster, J. Hai, Q. Le, G. Luxton, D. Martin, and J. Poen, "Image-guided radiosurgery for the spine and pancreas," *Comput. Aided Surg.*, vol. 5, pp. 278–288, 2000.

4. S. I. Ryu, S. D. Chang, D. H. Kim, M. J. Murphy, Q.-T. Le, D. P. Martin, and J. R. Adler, Jr., "Image-guided hypo-fractionated stereotactic radiosurgery to spinal lesions," *Neurosurgery*, vol. 49, pp. 838–846, 2001.
5. L. Lemieux, R. Jagoe, D. R. Fish, N. D. Kitchen, and D. G. T. Thomas, "A patient-to-computed-tomography image registration method based on digitally reconstructed radiographs," *Med. Phys.*, vol. 21, pp. 1749–1760, 1994.
6. G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, and D. J. Hawkes, "A comparison of similarity measures for use in 2D-3D medical image registration," *IEEE Trans. Med. Imaging*, vol. 17, pp. 586–595, 1998.
7. G. P. Penney, P. G. Batchelor, D. L. G. Hill, and D. J. Hawkes, "Validation of a two- to three-dimensional registration algorithm for aligning preoperative CT images and intraoperative fluoroscopy images," *Med. Phys.*, vol. 28, pp. 1024–1032, 2001.
8. R. A. McLaughlin, J. Hipwell, D. J. Hawkes, J. A. Noble, J. V. Byrne, and T. C. S. Cox, "A comparison of 2D-3D intensity-based registration and feature-based registration for neurointerventions," in *Medical Imaging Computing and Computer-Assisted Intervention (MICCAI) 2002*, T. Dohi and R. Kikinis, Eds. Berlin: Springer-Verlag, 2002, pp. 517–524.
9. J. R. Adler, Jr., M. J. Murphy, S. D. Chang, and S. L. Hancock, "Image-guided robotic radiosurgery," *Neurosurgery*, vol. 44, pp. 1299–1307, 1999.
10. J. B. West, J. M. Fitzpatrick, M. Y. Wang, B. M. Dawant, C. R. Maurer, Jr., R. M. Kessler, R. J. Maciunas, *et al.*, "Comparison and evaluation of retrospective intermodality image registration techniques," *J. Comput. Assist. Tomogr.*, vol. 21, pp. 554–566, 1997.
11. D. B. Russakoff, T. Rohlfing, D. Rueckert, R. Shahidi, D. Kim, and C. R. Maurer, Jr., "Fast calculation of digitally reconstructed radiographs using light fields," *Medical Imaging 2003: Image Processing*, vol. Proc. SPIE 5032, 2003 (in press).
12. C. Studholme, D. L. G. Hill, and D. J. Hawkes, "Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures," *Med. Phys.*, vol. 24, pp. 25–35, 1997.
13. T. Rohlfing, D. B. Russakoff, M. J. Murphy, and C. R. Maurer, Jr., "An intensity-based registration algorithm for probabilistic images and its application for 2-D to 3-D image registration," *Medical Imaging 2002: Image Processing*, vol. Proc. SPIE 4684, pp. 581–591, 2002.

Correction of C-Arm Projection Matrices by 3D-2D Rigid Registration of CT-Images Using Mutual Information

Ulrich Müller, Stefan Bruck, Jürgen Hesser, and Reinhard Männer

Institute for Computational Medicine
Universities Mannheim and Heidelberg
B6, 23, D-68131 Mannheim, Germany
ulrich.mueller@ti.uni-mannheim.de

Abstract. The paper discusses rigid registration of a 3D volume to 2D projections based on Mutual-Information similarity measures. The experiments we conducted are similar to those of Zoellei. However, in order to obtain a stopping criterion we use the Resilient Backpropagation as optimizer. Performing the registration on artificial images we obtain a speedup of approx. a factor of 10. We applied the matching approach to calibrate C-arm positions and found lateral movements of the detector of 0.5-1 mm in a given set of projections.

1 Introduction

The growing relevance of minimal invasive techniques in surgery and interventional diagnosis has lead to an increasing demand for medical imaging systems like CT, MRI etc. Pre-operative 3D images are widely used in intervention planning. In some cases it is necessary to follow deformations or movements of the patient during the operation or to visualize the location of a tool (e.g. a catheter tip). Acquiring 3D images intra-operatively is not yet a reasonable option because this would require a costly CT or MRI in or near the operation theatre.

Hence, X-ray fluoroscopy 2D images are used for intra-operative imaging. For combining the information of pre- and intra-operative images, 3D-2D registration is necessary. In order to do this properly, at least two 2D images are needed. These images are compared with virtual projections (DRRs) from the 3D images.

A special application of this matching process is the correction of the projection matrices of the C-arms in order to reconstruct high resolution volumes.

Therefore we investigate the alignment of a 3D volume reconstructed from CT-images acquired by a C-arm system and two X-ray fluoroscopic 2D images using rigid transformations. No fiducial markers are used. The algorithms we use are based on the MI method proposed by P. Viola [12].

1.1 State of the Art

Our approach of matching a 3D CT volume with 2D radiographs requires a measure of similarity that is used as target function for the optimization of the free parameters for matching.

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 161–170, 2003.

In this study we do not discuss feature-based approaches using e.g. fiducial markers or anatomical landmarks [1,2,3]. They require an extra camera that captures the scene and there is the problem of having markers fixed relatively to the patient.

Image based matching approaches have been proposed by several working groups. For example, Penney et al. [4,5] investigated this case using gradient based optimization and multiresolution. They achieved the best results using pattern intensity and gradient difference. Chung et al. [6] propose using a priori knowledge of the expected joint intensity distribution estimated from aligned training images. They use Kullback-Leibler Distance as similarity measure. Jonic et al. [7] use precomputed rescaled versions of both volume and images. They rely on a mean-squares cost function and a Levenberg-Marquardt algorithm for optimization. The achieved accuracy is 2.0 ± 1.0 mm on a level of a volume of 64×64×72 and images of 96×72. The registration takes about 54 sec on the lowest level and 152 sec at the next highest levels. The tests were done with images of a spine. LaRose [8] uses a variance-weighted sum of local normalized correlation (VLNC) as similarity measure and a quasi-Newton method for optimization. As a software approach, he proposes pre-computing a 4D-LUT called a transgraph which contains the rendering results for possible rays through the volume. This speeds up computing the DRRs. He also proposes hardware accelerated accumulation in texture mapping for volume rendering. Rohlfing et al. [10,11] propose a probabilistic similarity measure based on intensity distributions along simulated rays through the 3D image rather than ray sums, resulting in probabilistic DRRs. L. Zoellei [13] uses the approach described by P. Viola [12]. For probability density estimation, Parzen Windowing and Histogramming are used. As optimzation methods serve Powell's method and gradient ascent. Additionally, a multiresolution approach is applied.

In this paper, we present an algorithm that is based on the work of Viola and Zoellei. We use MI with Parzen Windowing as similarity measure. In contrast Resilient Backpropagation is selected as optimization method. By pre-segmenting the target 2D images with a simple thresholding method we can speed up the matching process and introduce an automatic stopping criterion for the optimization. In addition, multiresolution is applied. We show in this paper that our method with a decreased sample size for the algorithm will decrease the capture range, compared to the Zoellei experiments, while being much faster.

2 Transformation Description

We assume that the biplanar imaging coordinate system (i.e. the geometric relation of the two image planes) and the internal camera parameters are known. In our experiment, we use two given projection matrices P_1 and P_2 as a base. The two image planes are nearly rectangular to each other. Each (3×4) projection matrix P can be decomposed into

$$P = Int \circ D_d \circ R \circ D_c. \tag{1}$$

D_c is constant translation that provides that the rotation is performed around the object center. R is a rotation that we specify using unit quaternions (3 parameters). D_d is a translation that positions the object origin in the image coordinate system (3 parameters). Int is the internal matrix of the projection. It contains 5 internal camera parameters. We specify an initial 3D-Transformation using $R_i = Rot(0.5, 0.5, 0.5)$ which is a rotation of 60 degrees around the axis (1,1,1) and D_{di} as identity. This is useful for having the starting quaternions away from the boundaries.

Now we describe the optimization parameters as T, which consists of 3 quaternion modifiers q_0, q_1, q_2 and 3 translation modifiers t_x, t_y, t_z. In our experiments we specify the identity ($T = 0$) to be the optimal solution. Given $P_1 = Int \circ D_d \circ R \circ D_c$, and $T = (q_0, q_1, q_2, t_x, t_y, t_z)$, we get a new projection matrix

$$P_1(T) = Int \circ D_d \circ R \circ R_i^{-1} \circ D_{di}^{-1} \circ D'_{di} \circ R'_i \circ D_c \tag{2}$$

$$R'_i = Rot(0.5 + q_0, 0.5 + q_1, 0.5 + q_2)\ ;\ D'_d = D(t_x, t_y, t_z) \tag{3}$$

This way $P_1(0) = P_1$.

In each iteration step, there is a comparison between the DRR using P_1 and the DRR using $P_1(T)$. The same applies for P_2. We denote the DRRs with the names of their canonically assigned matrices.

3 Target Function: Mutual Information

The target function S for the images P_1, P_2 and the parameter set T is the mutual information I of both images, i.e.

$$S = I(P_1, P_1(T)) + I(P_2, P_2(T)) \tag{4}$$

The goal is to find T that maximizes S. This means the optimum $T = 0$. The Mutual Information I of two random variables A, B is hereby defined as

$$I(A, B) = H(A) + H(B) - H(A, B) \tag{5}$$

with $H(A)$ being the entropy of A, defined as the expected value of the negative log probability.

$$H(A) = E_A[-\log p(A)] \tag{6}$$

In each iteration step, we take a random sample a of image points. Then the probability distribution of the variables is estimated by a Parzen Window approach using Gaussian kernels. The probability density estimate of a random variable x for this sample is

$$P^*(x, a) = \frac{1}{n} \sum_{x_a \in a} g_\psi(x - x_a) \tag{7}$$

with the d-dimensional Gauss kernel defined as

$$g_\psi(z) = \frac{1}{(2\pi)^{\frac{d}{2}} |\psi|^{\frac{1}{2}}} \exp(-\frac{1}{2} z^T \psi^{-1} z). \tag{8}$$

In these equations, n is the number of points in the sample collection a and g is the Gaussian kernel using ψ as covariance matrix. For computing the entropy of this estimate, we assume a sample of all point pairs in $a \times a$, without the diagonal pairs.

$$h^*(x) = \frac{-1}{n} \sum_{x_b \in a} \log(P^*(x_b, a)) = \sum_{x_b \in a} \sum_{\substack{x_a \in a \\ x_a \neq x_b}} g_\psi(x_b - x_a) \tag{9}$$

4 Optimization: Gradient Ascent Using Resilient Backpropagation

We use gradient ascent for finding the maximum of our target function. In principle this is just a local method. But the target function is being smoothed by the stochastic approach described above.

Before optimizing, the target images P_1 and P_2 are segmented by a simple gradient thresholding method. This provides us with a set of "interesting" pixel positions. We denote this set as Ω_x.

4.1 Computing the Gradient

For the sample point x_i in the image P_1, we denote

$$u_i = P_1(x_i), v_i = (P_1(T))(x_i). \tag{10}$$

u_i is the gray value in image P_1. v_i is rendered by creating a single ray through the volume. This way, no full DRR $P_1(T)$ needs to be computed.

In each iteration step, we have to compute the gradient

$$\frac{d}{dT} I(u(x), v(T(x))) = \frac{d}{dT} h(u(x)) + \frac{d}{dT} h(v(T(x))) - \frac{d}{dT} h(u(x), v(T(x))) \tag{11}$$

$h(u(x))$ is not a dependent of T, so the first term vanishes. We estimate the gradient by

$$\frac{dI}{dT} \approx \sum_{x_i \in a} \sum_{\substack{x_j \in a \\ x_j \neq x_i}} (v_i - v_j)^T (W_v(v_i, v_j)\psi_v^{-1} - W_{uv}(w_i, w_j)\psi_{vv}^{-1}) \frac{d}{dT}(v_i - v_j) \tag{12}$$

with the notations

$$W_v(v_i, v_j) = \frac{g_{\psi_v}(v_i - v_j)}{\sum_{x_k \in a} g_{\psi_v}(v_i - v_k)}, \tag{13}$$

$$W_{uv}(w_i, w_j) = \frac{g_{\psi_{uv}}(w_i - w_j)}{\sum_{x_k \in a} g_{\psi_{uv}}(w_i - w_k)}, \tag{14}$$

$$w_i = [u_i, v_i]^T,\ w_j = [u_j, v_j]^T,\ \text{and}\ w_k = [u_k, v_k]^T \tag{15}$$

with ψ_v, ψ_{uu}, ψ_{vv} being scalars denoting the covariances and

$$\psi_{uv} = \mathrm{diag}(\psi_{uu}, \psi_{vv}). \tag{16}$$

For the approximation of the terms dv_i/dT we create an additional ray for each parameter of T and compute the difference quotients.

4.2 Resilient Backpropagation

Gradient ascent is a deterministic iterative optimization. In each step, the transformation T is updated. $T^{(i+1)} = T^{(i)} + \Delta T^{(i)}$, with the step size $\Delta T^{(i)}$ being determined by the gradient $\frac{dS}{dT}$. Resilient Backpropagation (Rprop) uses only the signs of the gradient components, not their absolute value.

$$\Delta T_j^{(i)} = \begin{cases} -\lambda_j^{(i)} & \text{if } \frac{\partial S}{\partial T_j}(T^{(i)}) < 0 \\ +\lambda_j^{(i)} & \text{if } \frac{\partial S}{\partial T_j}(T^{(i)}) > 0 \\ 0 & \text{if } \frac{\partial S}{\partial T_j}(T^{(i)}) = 0 \end{cases} \tag{17}$$

The step size is changed for each component, using two constant factors $0 < \eta^- < 1 < \eta^+$ for decelerating/accelerating.

$$\lambda_j^{(i)} = \begin{cases} \eta^+ \cdot \lambda_j^{(i-1)} & \text{if } \frac{\partial S}{\partial T_j}(T^{(i-1)}) \frac{\partial S}{\partial T_j}(T^{(i)}) > 0 \\ \eta^- \cdot \lambda_j^{(i-1)} & \text{if } \frac{\partial S}{\partial T_j}(T^{(i-1)}) \frac{\partial S}{\partial T_j}(T^{(i)}) \leq 0 \end{cases} \tag{18}$$

The advantage of this method is that the influence of outliers is minimized in the computation.

The iteration stops when each $\lambda_j^{(i)}$ falls below a minimum learning rate. In fact, only two learning rates were used, one for the quaternions, one for the translations.

5 Multiresolution

In our experiments we have observed that scaling down the volume and the target images is useful in terms of speed and robustness. The set Ω_x of points from which the samples are drawn is decreased without becoming less representative. This increases the capture range of the translation parameters by a factor of 1.8 per stage, due to the bigger pixel size. Working on smaller data sets decreases accuracy. We do the registration on the smallest level and take the result as starting point for the next higher level. Since $T^{(0)}$ being close to the optimum now, the learn rate can be chosen to be smaller. In a bigger dataset, the pixels and voxels cover smaller areas, hence a higher accuracy is obtained.

6 Experiments

6.1 Test Description

We work with a skull phantom filling a cube of 256^3 voxels, reconstructed from a set of 100 projections of size 1024^2 and their corresponding projection matrices. We take two projection matrices P_1 and P_2 that project this cube onto

image planes which are nearly rectangular to each other. As described before, $T = 0$ is the optimal solution. By ray-casting with space-leaping and early-ray-termination we render images of size 256^2. Then we change to a smaller level. The volume is downsized to a 128-voxel cube and two images of size 128^2 are rendered using the matrices P_1 and P_2 changed accordingly. The smallest level is size 64.

We start the experiments at level size 64. We have a volume and two images with their matrices P_1 and P_2. We choose an angle r_0 and a translation distance d_0. $T^{(0)}$ is a rotation of r_0 around a random axis and a translation which is chosen as a random point on the sphere with radius d_0 (i.e. the translation vector has length d_0).

Fig. 1. First target image.

Fig. 2. Second target image.

We use the VGL library [17] for our rendering. This way we can use fast rendering techniques like space leaping for our purposes. Rendering a sample of 20 points from a 64-volume takes roughly 3.8 ms. We have observed that the rendering takes about 90% of the time for the algorithm.

For computing the mutual information we choose to segment the images (normalized to range $[0, 1]$) with a gradient threshold of 0.035 to get a Ω_x of size 400-600 for each image. We use a sample size of 20. So, for each image 20 * (1+6) rays (one for computing $v(T(x))$, six for computing dv_i/dT) need to be computed, which means 280 rays per iteration. We use an iteration maximum of 500. When this value is reached for the third time without convergence, the result is a miss.

The resulting T gives an error rotation and an error translation. We denote the rotation with its angle r_e independent of its axis. d_e is the length of the error translation vector. We consider the result as being good if $r_e < 1$ degree and $d_e < 1$ voxel unit. With a 64×64 image, one voxel has the size of about 4mm.

The diagrams show level 64 experiment results for the capture range. We choose starting positions of $d_0 = 0$ and variable r_0 resp. $r_0 = 0$ and variable d_0. We find that producing a miss takes considerably more time than finding a good result. Achieving a good result is faster when the initial distance to the optimum is smaller.

Between the first two rows of the table, we have decreased the minimum learning rate for the optimization. As we see, this improves the accuracy while

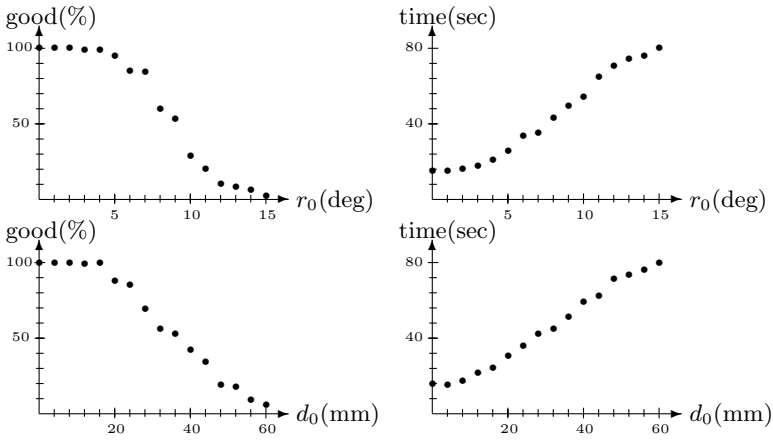


Fig. 3. Top Left: Success rate as a function of the start angle, Top Right: Time consumed for that angle, Bottom: Same as above for translation as parameter.

Table 1. Experiments showing success rate, average μ and standard deviation σ of rotation and translation error, and average time consumption of registration.

Level	$r_0(deg)$	$d_0(mm)$	good (%)	r_e (μ)	r_e (σ)	d_e (μ)	d_e (σ)	time (sec)
64	5	12	92.6	0.4865	0.2064	0.7196	0.3489	24.7
64	5	12	94.2	0.4054	0.1867	0.5784	0.2716	30.9
64	5	16	88.4	0.3804	0.1756	0.7204	0.2904	37.2
64	6	12	85.4	0.3726	0.1754	0.5168	0.2624	39.6
128	1	4	100	0.1780	0.0794	0.3272	0.1782	5.9

taking more time. In rows 2-4, only the initial $T^{(0)}$ changes. Like above, we observe changes in robustness and time consumption. For the level 128 we consider starting with a good result from the 64 level. Unlike level 64, we do the optimization only once, because no check is needed. For level 256 we have not observed any improvement in accuracy, so we omit the results here.

We want to compare our results with the experiments of Zoellei [13]. From that thesis we only use the Gage skull experiments in which all 6 transformation parameters are optimized simultaneously and Parzen windows are used for computing the mutual information measure. The Gage skull is downsized two times. We denote the 256×256×212 volume as $sm1$ and the 128×128×106 volume as $sm2$. X-ray images serve as target images that have been registered by hand to the virtual images in order to find the ideal transformation. The starting values are randomly picked in the specified ranges. Time and accuracy are observed in different experiments.

The main differences of Zoellei's experiments (in the first 4 rows) to ours are as follows: Zoellei used a steepest descent algorithm that stops after a given iteration maximum which is 5000 for this experiment. Our learn rate starts with 0.002/0.8 for rotation/translation and terminates when all parameters fall below

Table 2. A comparison of the Zoellei experiments to ours.

Level	sample	$r_0(deg)$	$d_0(mm)$	good (%)	$r_e(deg)$	$d_e(mm)$	time (sec)
sm1	100-150	0-20	0-17.3				6337.3
sm2	100-150	0-20	0-17.3				3567.3
sm1	100-150	0-15	0-14.1	90	0.2256	2.4073	
sm2	100-150	0-15	0-14.1	90	0.4835	4.1094	
64	20	0-20	0-17.3	37	0.3537	0.6284	56.6
64	20	0-10	0-17.3	76	0.3873	0.6852	34.5
128	20	1	4	100	0.1780	0.3272	5.9
256	100	3	3	78	0.4494	0.5452	121.9
256	80	3	3	78	0.4613	0.5607	108.4
256	60	3	3	70	0.5372	0.6367	92.7

0.00036/0.16, thus reducing the amount of iterations to 200-800, depending on the initial error. The other big difference lies in the data sets. Our target images are identical to the optimal DRRs. This results in a higher accuracy. We only need a sample size of 20, compared to 100-150 of Zoellei. Hence for one iteration, less rays have to be rendered. Both effects lead to our advantage in speed, while decreasing the capture range.

The last three rows show the experiments with the original images, downsized to 256×256, with changing sample sizes. We have used corrected projection matrices (see next section) and the corresponding original images as targets. We observe a tradeoff of time and robustness/accuracy for all parameters.

7 Using Registration for Correcting Geometry in the 3D-Reconstruction from C-Arm CT Projections

During one 3D-image acquisition run, a C-arm X-ray-device produces a set of images along with their projection matrices containing the geometrical parameters of the state of the C-arm during the acquisition. Then the 3D volume is reconstructed from this data. Due to the mechanical construction of the C-arms there are deviations between the real positions and those that measured by the system. We introduce an iterative method for improving the reconstruction by only using the existing data of the patient acquisition [16].

From the acquisition, a set of images and corresponding projection matrices is given. Each matrix consists of 5 internal and 6 external parameters. Our geometry correction only considers the external parameters. The iteration works as follows:

1. Reconstruct the volume from the given images using the matrices.
2. Compute virtual projections (DRRs) from the reconstructed volume using the given matrices.
3. Compute the distance between each given and virtual projection and determine the number i of the worst virtual projection.

4. Do a 3D-2D registration using the volume and the given image i. The result is a new projection matrix.
5. Take the new matrix and go back to step 1.

During one iteration one matrix and the volume are updated. The given images remain the same. The iteration terminates when the worst deviation in step 3 is lower than a given threshold or a maximum of iteration steps is reached.

The reconstruction in step 1 is done using the Feldkamp-Algorithm [15]. In step 3 we use squared distance correlation as distance measure. Step 4 contains the matching process. Given is a reconstructed 3D volume, a 2D image and a projection matrix. The external parameters of the matrix produce the 3D-transformation T to be optimized. The optimization is done as described in sections 3-5.

For our experiments, we use the same data set as in the last section. When we register the reconstructed volume to the original images, we cannot assume proper knowledge about the geometry between the image planes (because the matrices are not exact), so we can only use one image at a time for calibration. During 3D-2D registration only the translation in x- and y-direction and rotation around the z-axis (parallel to the projection direction) have shown convergence. For this reason, a 2D-2D approach that uses only these 3 parameters is better suited. We compute a complete DRR with the given matrix and apply a 2D-2D matching to the original target. This way we can use the DRR for choosing our point set Ω_x. In the DRR the gradient threshold method is undisturbed by any shadows that might be in the original image. We have applied this method to a set of 25 out of 100 images that are used to reconstruct our volume. For each image, we have started with $T = 0$ and done the matching 100 times on the 256-level i.e. the downsized images. For each image, we have recorded the T found. We then have computed the average error and the standard deviation for all 25 images. We did the same experiment with 13 images on the 1024-level using a sample size of 100.

Table 3. Experiments showing average μ and standard deviation σ of observed rotation and translation error in degree resp. mm, and average time for one matching.

Level	z-Rot (μ)	z-Rot (σ)	x-Tra (μ)	x-Tra (σ)	y-Tra (μ)	y-Tra (σ)	time (sec)
256	0.1847	0.2371	0.6750	0.1234	0.6418	0.2401	3.3
1024	0.1636	0.2154	0.7066	0.0523	0.5622	0.0411	10.8

The standard deviation shows the accuracy of all experiments converging to the same MI maximum. We have found that regarding translation, we are able to find an error in the given projection matrices with an acceptable accuracy. For the rotation the proposed errors are smaller than our accuracy allows for.

We have done a 2D-2D matching on all 100 images of level 1024 and have obtained new projection matrices for the given images. Reconstructing with the new matrices yields a volume that is displaced by roughly 0.5mm in comparison to the old volume and is more accurate (the vessel contours are sharper).

References

1. J. Weese, T.M. Buzug, et al. An approach to 2D/3D Registration of a Vertebra in 2D X-ray Fluoroscopies with 3D CT images. CVRMed-MRCAS (1997) Grenoble.
2. T.S.Y. Tang, R.E. Ellis, G. Fichtinger. Fiducial Registration from a Single X-ray Image; A New Technique for Fluoroscopic Guidance and Radiotherapy. MICCAI (2000), pp. 502-511.
3. A. Liu, E. Bullit, S.M. Pizer. 3D/2D Registration via skeletal near projective invariance in tubular objects. MICCAI (1998), pp. 952-963.
4. G. Penney, J.Weese et al. A Comparison of Similarity Measures for Use in 2D-3D Medical Image Registration. MICCAI (1998), pp. 1153-1161.
5. J. Hipwell, G. Penney et al. 2D-3D Intensity Based Registration of DSA and MRA - A comparison of similarity measures. MICCAI (2002), pp. 501-508.
6. A. Chung, W. Wells et al. Multi-modal Image Registration by Minimising Kullback-Leibler Distance. MICCAI (2002), pp. 525-532.
7. S. Jonic, P. Thevenaz, M. Unser. Multiresolution-based registration of a volume to a set of its projections. Swiss Federal Institute of Technology, Lausanne, 2001(?).
8. D. LaRose. Iterative X-ray/CT Registration using Accelerated Volume Rendering. PhD-Thesis Carnegie Mellon University, 2001.
9. T. Rohlfing, J. Beier et al. Multimodale Registrierung mit effizienten Lernverfahren für neuronale Netze. Lehmann et al.(Hrsg.), Bildverarbeitung für die Medizin 1998: Algorithmen-Systeme-Anwendungen, RWTH Aachen, Springer.
10. T. Rohlfing, D.B. Russakov et al. An intensity-based registration algorithm for probabilistic images and its application for 2-D to 3-D registration. vol. 4684 of Proceedings of SPIE, 2002.
11. T, Rohlfing, C. R. Maurer. A Novel Image Similarity Measure for Registration if 3-D MR Images and X-Ray Projection Images. MICCAI(2002), pp. 469-476.
12. P. Viola. Alignment by Maximization of Mutual Information. PhD-Thesis, Massachusetts Institute of Technology, 1995.
13. L. Zoellei, 2D-3D Rigid-Body Registration of X-Ray Fluoroscopy and CT Images. Masters Thesis, Massachusetts Institute of Technology, 2001.
14. Xu Meihe, R. Srinivasan, W.L. Nowinski. A Fast Mutual Information Method for Multi-modal Registration. LNCS 1613, pp 466-471, Springer Verlag
15. L. A. Feldkamp, L.C Davis, J.W. Kress, Practical cone-beam algorithm, J. Opt. Soc. Amer., A/Vol 1 Nr.6, 612-619, 1984.
16. S. Bruck, Geometriekorrektur bei der 3D-Rekonstruktion aus C-Arm-Projektionen. Masters Thesis, Universität Mannheim, 2002.
17. www.volumegraphics.com

Image Registration for Distortion Correction in Diffusion Tensor Imaging

Thomas Netsch[1] and Arianne van Muiswinkel[2]

[1] Philips Research Laboratories
Roentgenstr. 24–26, 22335 Hamburg, Germany
thomas.netsch@philips.com
[2] Philips Medical Systems
Veenpluis 6, 5680 DA Best, The Netherlands

Abstract. In diffusion tensor imaging the calculation of functional information is limited by head movement and eddy current-induced image distortion. In this paper the application of image registration for distortion correction is investigated. In particular, a 3D affine and a dedicated transformation which is adapted to the type of distortion and the similarity measures mutual information and local correlation are compared to each other. The registration results are quantitatively evaluated by analyzing their consistency properties. Visual inspection shows that registration generally improves the quality of the functional information. The consistency tests reveal that both transformations provide similar registration results which is remarkable since the dedicated transformation does not take advantage of modeling of the underlying imaging physics. Furthermore, it is shown that local correlation similarity is an interesting alternative to mutual information. The registration of a DTI series with local correlation is more consistent and takes only about one minute for calculation.

1 Introduction

In magnetic resonance (MR) imaging, diffusion tensor imaging (DTI) offers the possibility of mapping properties closely related to functional architecture of the brain in vivo [1]. Generally, water diffuses more easily along axon bundles and therefore the anisotropy of the diffusion process gives information about their shape and direction. For the calculation of the diffusion tensor, at least 6 distinct measurements with non-parallel diffusion-weighted gradients and one reference measurement without gradient are required, see Fig. 1. For each voxel this information is combined to yield an estimate of the tensor matrix. Based on the eigenvalues of the diffusion tensor, parameter maps such as fractional anisotropy (FA) can be derived. More advanced representations are connectivity maps visualizing similarities of the diffusion tensor in adjacent voxels or color-coded field maps [2].

Single-shot echo-planar imaging (EPI) is the most common read-out method for DTI. Unfortunately, this fast acquisition scheme is highly sensitive to eddy

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 171–180, 2003.

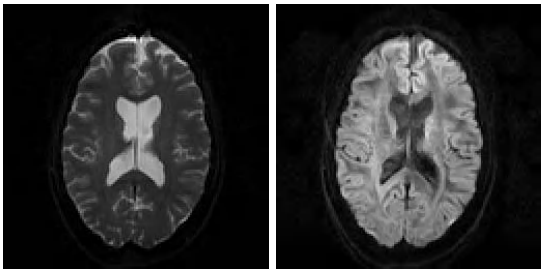


Fig. 1. Corresponding slices of the reference image of the acquisition (left) and one of the 6 diffusion-weighted images (right) of a standard 3 T DTI series.

current-induced gradient transients that arise from switching the strong diffusion-weighting gradients which may lead to shear, scaling and shift of the acquired slices as shown in Fig. 2. Since the diffusion tensor calculation relies on the assumption that a given voxel corresponds to the same anatomical location in all measurements, eddy currents—as well as patient movement during the acquisition—cause distortions of several millimeters in FA maps [3].

Apart from methods improving the diffusion pulse sequence the retrospective application of image registration is frequently suggested for distortion correction. However, differences in image contrast due to dissimilar diffusion properties of the brain tissue only allow the use of similarity measures suitable for multi-modality registration such as mutual information (MI) [4,5,6]. To avoid such difficulties, Andersson and Skare [7] directly minimize the residual error of the diffusion tensor calculation by simultaneously registering all images of the diffusion series. Unfortunately, this is only possible if more than 6 measurement, preferably, 12–20 different gradient directions are available. For the eddy current distortion correction a dedicated transformation derived from MR imaging physics is suggested [3]. As proposed by Andersson and Skare [6], this transformation can be combined with a rigid transformation to address the correction of head motion as well.

Most of current work, however, only shows feasibility of the proposed correction or compares the results by visual inspection of parameter maps since ground truth data is not available in DTI. Quantitative evaluation becomes even more important when taking into account that already a general 3D affine registration [8] can significantly improve the quality of FA maps as illustrated in Fig. 3. As a result, transformations which are even more adapted to the physical model of the distortion could further improve the quality of parameter maps.

In this contribution we investigate current work on DTI registration. In particular, we compare a 3D affine transformation and a dedicated type of image transformation which addresses both correction of head motion and eddy current distortion simultaneously [7]. For quantitative comparison the consistency properties of the registration results are analyzed [9] and further augmented by a novel concept of cross consistency testing. In medical imaging only MI has gained widespread acceptance for multi-modality registration problems. How-

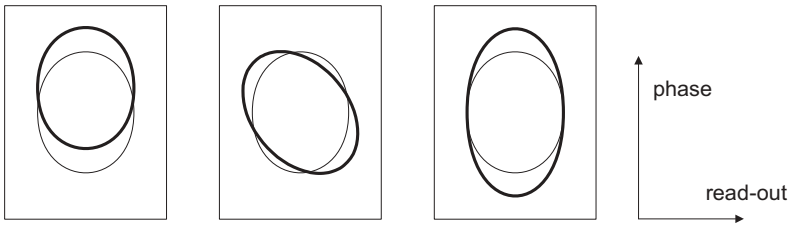


Fig. 2. Symbolic representation of eddy current-induced geometric distortion of a slice of diffusion-weighted image (shift, shear and scaling, from left to right) [3].

ever, we have made good experience with local correlation (LC) similarity [10] in the case of MR-CT matching and the alignment of brain perfusion time-series [8]. Therefore, we compare MI to the registration results of LC. In the next section we describe the dedicated distortion correction transformation and introduce LC and its optimization.

2 Method

Grey value-based registration methods spatially align a target image $\mathbf{t}$ to a reference image $\mathbf{r}$ with respect to a transformation $\mathbf{T}$ described by parameters $\mathbf{p}$. We denote the grey value of $\mathbf{r}$ at voxel $\mathbf{x}_i$ as $\mathbf{r}|_{\mathbf{x}_i}$ or r_i for short. During the registration the similarity of the grey value $\mathbf{r}|_{\mathbf{x}_i}$ is compared to the grey value at the corresponding position of the target $\mathbf{t}$,

$$\mathbf{t}|_{\mathbf{T}^{-1}(\mathbf{p})\ \mathbf{x}_i} = \mathbf{t}|_{\mathbf{x}'}\ , \tag{1}$$

which requires the inverse transformation $\mathbf{T}^{-1}$ and image interpolation since $\mathbf{x}'$ is usually not a grid point of $\mathbf{t}$. For a given $\mathbf{p}$, the quality of the match between $\mathbf{r}$ and the transformed target $\mathbf{t}$ is determined by a similarity function

$$\mathcal{M}(\mathbf{p}) = \sum_i F(\mathbf{r}|_{\mathbf{x}_i}, \mathbf{t}|_{\mathbf{T}^{-1}(\mathbf{p})\ \mathbf{x}_i}) \tag{2}$$

where F denotes a functional relationship of the voxels of both images.

2.1 Image Transformation

Assuming the distortion model by Jezzard *et al.* [3] and rigid-body movement of the subject, we follow Andersson and Skare [7] and use the following combined correction transformation

$$\mathbf{x} = \mathbf{T}(\mathbf{p})\ \mathbf{x}' = \mathbf{R}(\mathbf{p})\ \mathbf{D}(\mathbf{p})\ \mathbf{x}' \tag{3}$$

mapping a voxel $\mathbf{x}'$ from the distorted diffusion-weighted image space to the undistorted space of the reference image. $\mathbf{R}(\mathbf{p})$ is a 3D rigid transformation

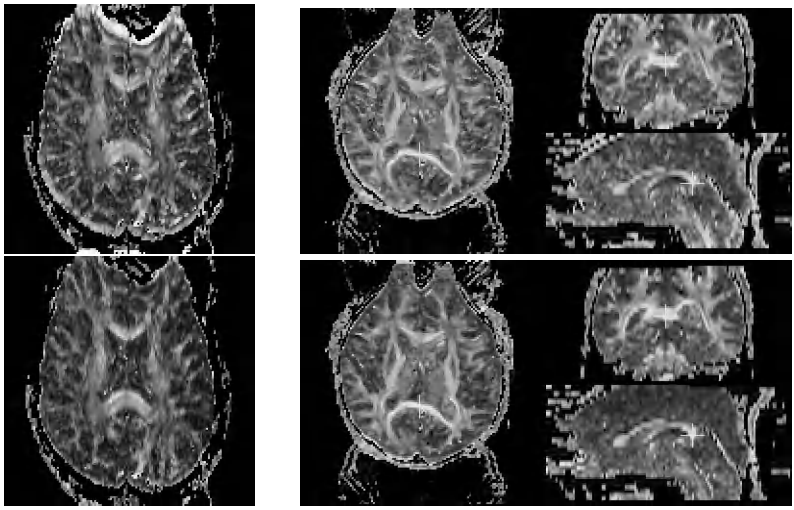


Fig. 3. Corresponding slices of FA maps (left) and ortho-views of the FA map (right) of the original data (top) and after its registration with 3D affine transformation (bottom). The effect of registration is visible at the edge of the brain (left) and in the sagittal view of the FA map (right): the structure of the corpus callosum and corona radiata is better defined after registration.

matrix in homogeneous coordinates with parameters p_0, p_1, p_2 defining rotation and p_3, p_4, p_5 translation. The distortion correction matrix is

$$\mathbf{D}(\mathbf{p}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \alpha(\mathbf{p}, z') & \beta(\mathbf{p}, z') & 0 & \gamma(\mathbf{p}, z') \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4}$$

where the α, β and γ specify shear, scaling and shift depending on $\mathbf{p}$ and slice z'. The matrix $\mathbf{D}$ assumes that the phase encoding direction coincides with the y-direction of the images. Since each slice may have an individual distortion we need to determine as many parameters $\mathbf{p}$ as slices are acquired to completely correct the entire volume. Assuming a smooth transition of shear, scaling and shift in the direction of the slices we further constrain the parameter space by considering only solutions that can be described by first order polynomials, hence $\alpha(\mathbf{p}, z') = p_6 z' + p_7$, $\beta(\mathbf{p}, z') = p_8 z' + p_9$ and $\gamma(\mathbf{p}, z') = p_{10} z'$. Note, the translation in the y-direction of the rigid part $\mathbf{R}$ already corrects for the constant y-shift of $\mathbf{D}$.

The transformation in Eq. (3) describes the mapping of a diffusion-weighted image to the reference image of the acquisition. The latter image is not affected by eddy currents since a diffusion gradient is not applied. If the acquisition reference is registered to a diffusion-weighted image, the transformation starts with $\mathbf{R}$ followed by a left-multiplication of $\mathbf{D}$.

A 3D affine transformation is also capable of simultaneously correcting rigid-body movement and zero order eddy current-induced distortion. A first order dependency of the slice number, however, is only addressed by the translational components of the transformation. In contrast to the diffusion transformation given by Eq. (3) there is one additional parameter. A 2D affine registration would allow the correction of the slices individually, but it is sensitive to registration outliers since the parameters of one slice are not connected to the parameters of the neighboring ones.

2.2 Local Correlation Similarity

Local correlation is based on the squared cross correlation coefficient

$$\mathrm{CC}(\mathbf{r},\mathbf{t}) = \frac{(\sum_i (r_i - \overline{\mathbf{r}})(t_i - \overline{\mathbf{t}}))^2}{\sum_i (r_i - \overline{\mathbf{r}})^2 \sum_i (t_i - \overline{\mathbf{t}})^2} \tag{5}$$

of $\mathbf{r}$ and $\mathbf{t}$ where $\overline{\mathbf{r}}$ and $\overline{\mathbf{t}}$ are the mean grey value of the images. Global CC is not suited for multi-modality registration since a global linear transfer of the grey values cannot be presumed. However, in a small neighborhood $n(j)$ of voxel j the assumption of a linear relationship is valid and the correlation coefficient indicates how similar the grey value transfers in both images are.

Therefore, the idea is to compute for a number of voxels j the squared cross correlation coefficient of the reference grey values r_i, $i \in n(j)$ and the resulting grey values t_i of the target and to accumulate the corresponding correlation values. The sum is defined as local correlation (LC) similarity of $\mathbf{r}$ and $\mathbf{t}$ [10]. Note, the squared coefficient guarantees that positive as well as negative correlated grey value transitions within $n(j)$ are considered as similar which is reason why LC can register also quite dissimilar images. We define for a given j

$$[\mathbf{uv}]_j \stackrel{\mathrm{def}}{=} \sum_{i \in n(j)} (u_i - \overline{\mathbf{u}_j})(v_i - \overline{\mathbf{v}_j}) \tag{6}$$

where $\overline{\mathbf{u}_j}$ and $\overline{\mathbf{v}_j}$ are the mean value of the neighborhoods. By the use of this notation, LC is defined as

$$\mathrm{LC}(\mathbf{r},\mathbf{t}) = \frac{1}{N} \sum_j \frac{[\mathbf{rt}]_j^2}{[\mathbf{rr}]_j [\mathbf{tt}]_j} \ . \tag{7}$$

The remaining question is which neighborhoods $n(j)$ are selected. LC should be large if the images are well-aligned. Obviously, neighborhoods placed at the edges of the reference image provide large correlation in the case of alignment since it can expected that the edges match corresponding ones in the target image. It can be shown [8] that it suffices to calculate the intensity variance for each neighborhood $n(j)$ and select those with the highest variance. The number of selected voxels with respect to the total number of image voxels is denoted as LC fraction. It is usually chosen between 5–20%. If the LC fraction

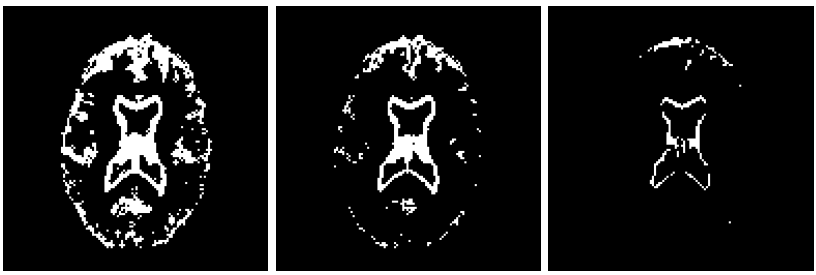


Fig. 4. Selected voxels for 10%, 5% and 1% LC fraction (left to right) shown for the slice of the reference image already depicted in Fig. 1 (left). and 1.

is large, neighborhoods in image areas with low variance values are selected resulting in many small correlation coefficients and hence slow convergence of the optimization. If the number of voxels is too small, registration may become inaccurate since the registration is based on too few image features. Note, smaller LC fractions generally result in less computation since only a part of the volume is considered for registration. Fig. 4 depicts the voxels selected in the reference image of the diffusion series already shown in Fig. 1 for different LC fractions.

2.3 Least-Squares Optimization of Local Correlation

A quadratic from of F in Eq. (2) allows the application of efficient least-squares optimizers such as Gauss-Newton methods.Gauss-Newton optimization requires only first order derivative information while achieving super-linear convergence rates. Apart from the quadratic from of F it must be guaranteed that the similarity function becomes small in the vicinity of the minimum. We therefore do not consider LC for optimization but $1 - \mathrm{LC}$ since local correlation coefficients tend to one if the images are aligned. It is

$$1 - \mathrm{LC} = \frac{1}{N} \sum_j \left(1 - \frac{[\mathbf{rt}]_j^2}{[\mathbf{rr}]_j [\mathbf{tt}]_j} \right) \tag{8}$$

$$= \frac{1}{N} \sum_j \left(\frac{1}{[\mathbf{rr}]_j} \left[\left(\mathbf{r} - \frac{[\mathbf{rt}]_j}{[\mathbf{tt}]_j} \mathbf{t} \right) \left(\mathbf{r} - \frac{[\mathbf{rt}]_j}{[\mathbf{tt}]_j} \mathbf{t} \right) \right]_j \right) \tag{9}$$

$$= \sum_j \sum_{i \in n(j)} \frac{1}{N [\mathbf{rr}]_j} \left((r_i - \overline{\mathbf{r}_j}) - \frac{[\mathbf{rt}]_j}{[\mathbf{tt}]_j} (t_i - \overline{\mathbf{t}_j}) \right)^2 \tag{10}$$

yielding the required expression in least-squares form [8].

For an efficient reformation of the target image during optimization an analytical form of the inverse transformation $\mathbf{T}^{-1}(\mathbf{p})$ is required. To completely avoid the inversion of Eq. (3) the reference and the target are interchanged in the very beginning of the registration. As a consequence, the Gauss-Newton optimization is built on the 'reference' image $\mathbf{t}|_{\mathbf{x}'_i}$, the reformatted 'target' image

$\mathbf{r}|_{\mathbf{T}(\mathbf{p})\ \mathbf{x}_i'}$ and its partial derivatives with respect to $\mathbf{p}$. The partial derivatives are calculated by the method of finite differences.

3 Consistency Testing

Consistency testing for the evaluation of registration results has already been used in medical image registration [9], in particular in the case of time-series registration [11]. The registration consistency for a series of images is calculated as follows. Given the images $\mathbf{h}_k$ and $\mathbf{h}_l$, two cyclic registrations are performed: $\mathbf{h}_k$ is registered to $\mathbf{h}_l$ and $\mathbf{h}_l$ is registered to the image $\mathbf{h}_k$ yielding the registration transformations $\mathbf{T}_{lk}$ and $\mathbf{T}_{kl}$, respectively. Obviously, their composition $\mathbf{T}_{lk}\mathbf{T}_{kl}$ can be related to the inherent consistency of the applied registration algorithm: it should be close to the identity. The consistency error θ_{kl}, $k \leq l$ is defined as

$$\theta_{kl} = \max_{\mathbf{x}_i} \|(\mathbf{T}_{lk}\mathbf{T}_{kl})\ \mathbf{x}_i - \mathbf{x}_i\|_2 \tag{11}$$

measuring the largest displacement given in [mm] space.

A time-series of N images allows the calculation of $\frac{1}{2}N(N+1)$ consistency errors θ_{kl}. Their root mean squared (RMS) value together with its standard deviation is denoted as the consistency error of the underlying registration algorithm. However, in the case of a standard DTI experiment with 6 different gradient directions only 6 errors $\theta_{01}, \ldots, \theta_{06}$ can be calculated due to the dedicated type of correction transformations defined by Eq. (3).

Note, the consistency error and the accuracy of a registration are different concepts of measurement. Consistency testing can only be related to the registration accuracy if the amount of 'misalignment' is not biased. If there is any systematic registration error, for example, because the algorithm always underestimates the correct transformation, then the individual errors may be highly correlated. The true registration error may be greater than the consistency-derived error estimate suggests.

To overcome such limitation we introduce a novel concept of cross consistency. Let $\mathbf{T}_{kl}$ and $\mathbf{S}_{kl}$ be the registration results of two algorithms, for example MI and LC registration, which are applied to a DTI series. In analogy to Eq. (11) the cross consistency error ϕ_{kl}, $k \leq l$ is defined as

$$\phi_{kl} = \max_{\mathbf{x}_i} \|(\mathbf{T}_{lk}\mathbf{S}_{kl})\ \mathbf{x}_i - \mathbf{x}_i\|_2\ . \tag{12}$$

The RMS value and its standard deviation measures the compatibility of the registration results of two algorithms. If the cross consistency error is of the same order as the consistency error of the individual results, then it can be concluded that both underlying registration algorithms are very similar with respect to their capabilities of the alignment of the images: the impact of exchanging registration results has only a minor influence on the consistency. A significantly larger cross consistency error compared to the individual errors indicates that the consistency error is very sensitive when registration results are

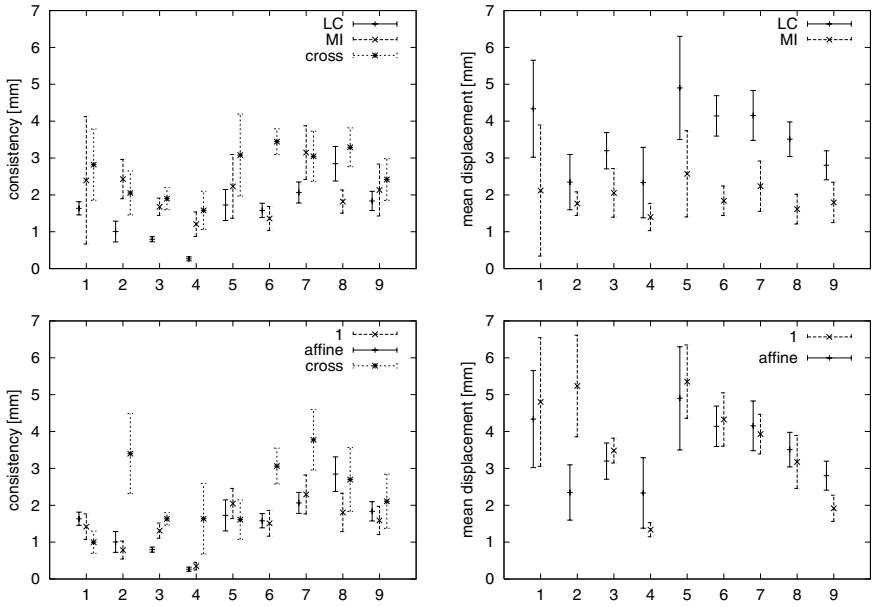


Fig. 5. Cross consistency tests comparing LC and MI in the case of affine registration (top) and diffusion and affine registration for LC (bottom) in the case of 9 DTI series.

interchanged. Therefore, it can be expected that the underlying algorithms also align the images of the time-series in a different, usually wrong manner.

The concept of cross consistency can also be extended to a measure of 'mean displacement' of the voxels when registering a DTI series. For example, a registration algorithm which always reports transformations close to identity is very consistent but might not accurately align the images. This situation can be identified if the mean displacement is small as well. The mean displacement can be defined as the cross consistency error by replacing $\mathbf{S}_{kl}$ in Eq. (12) by the identity transformation.

4 Results

The registration is tested with 9 standard DTI series acquired with a clinical 3 T (series **1**–**3**) and 1.5 T (series **4**–**9**) scanner (Gyroscan ACS-NT, Philips Medical Systems). Each series consists of 6 diffusion directions. The number of slices varies between 12 and 45 a with constant image matrix of (128×128). The slice direction is usually transversal, series **4** is coronal, **7** and **8** have sagittal direction. Phase direction corresponds to y-direction of the images. LC templates are determined for a $(3 \times 3 \times 3)$ neighborhood and a LC fraction of 10% is used. Image reformations during the optimization process use linear interpolation; for the resampling of the final result, however, a higher order B-spline interpola-

tion scheme is applied. The MI algorithm is implemented by a gradient descent optimization [12].

Fig. 5 compares LC and MI similarity in the case of affine registration by means of cross consistency tests (top). LC provides more consistent registrations results in general (left). Quantitatively, we denote two registrations as compatible if the error bar (mean ± standard deviation) of the cross consistency error overlaps at least one of the individual consistency error. With respect to this definition, LC and MI are compatible for all series except for series **6**, whereas the corresponding mean displacements of LC (right) are much larger than those for MI. Note, both observations—larger displacements in combination with smaller consistency error—do not necessarily indicate that LC provides better registration quality when compared to MI. For example, it might happen by error that LC consistently shifts the images a few millimeters which would result in similar graphs as shown in Fig. 5. If we further assume for a moment that MI registration would provide a gold standard in DTI registration then both result could not be compatible. However, both measures are compatible. Thus, LC is an interesting, more consistent alternative to MI.

Fig. 5 compares the combined diffusion correction given by Eq. (11) and 3D affine transformation (bottom). The cross consistency (left) shows that the transformations are compatible for series **1**, **3**, **5**, **8** and **9**. In these cases also the mean displacement of affine and diffusion transformations is quite similar. Therefore, it can be concluded, that both transformations provide comparable registration quality. The other series, **2**, **4**, **6** and **7**, are not compatible. In particular, series **2** and **4** also show a significant difference in the mean displacement. The latter two series are the most promising examples to find out whether there is any benefit from the dedicated type of correction transformations. The visual inspection of their FA maps shows that there is a slight difference but a significant advantage of one of the transformations is not apparent. In a forthcoming paper we will address these differences more thoroughly.

5 Discussion

In this contribution we investigated image registration methods for their suitability to correct for head motion and eddy current-induced image distortion. The quantitative assessment of the registration results is addressed by evaluating their statistical consistency. Specifically, we compared 3D affine and a dedicated type of correction and we compared MI and LC for a 3D affine transformation.

The visual inspection of FA maps shows that all registrations performed improve their quality when compared to the maps calculated for the original series. However, the consistency tests show for 5 of 9 DTI series used in evaluation that both transformations must provide very similar registration results. It is remarkable, that the dedicated transformation does not take more advantage of the specific modeling of the eddy current-induced distortions. Possible explanations are that the use of a first order polynomial to describe the dependency of the parameters on the slice z' is too restrictive or that the model does not completely

reflect the nature of the distortions. Note, if we do not allow a dependency on z', the resulting transformation is a subspace of affine transformations.

Registrations with LC similarity are an interesting alternative to MI since the results are more consistent and provide larger voxels displacements while being still compatible with MI. In particular, due to the least-squares optimization of LC, registration of a DTI series is on the order of only 1 minute. In contrast, the applied MI algorithm requires more than 1 hour computation time.

References

1. Rowley, H.A., Grant, P.E., Roberts, T.P.: Diffusion MR imaging. Neuroimaging Clinics of North America **9** (1999) 343–361
2. Jaermann, T., Crelier, G., Pruessmann, K.P., Golay, X., Netsch, T., van Muiswinkel, A., Mori, S., van Zijl, P.C., Kollias, S., Boesiger, P.: SENSE-DTI at 3 Tesla. Submitted to Magnetic Resonance in Medicine (2003)
3. Jezzard, P., Barnett, A.S., Pierpaoli, C.: Characterization of and correction fro eddy current artifacts in echo planar diffusion imaging. Magnetic Resonance in Medicine **39** (1998) 801–812
4. Bammer, R., Auer, M.: Correction of eddy-current induced image warping in diffusion-weighted single-shot EPI using constrained non-rigid mutual information image registration. In: Proceedings of the ISMRM, Glasgow (2001) 508
5. Mangin, J.F., Poupon, C., Clark, C., Le Bihan, D., Bloch, I.: Eddy-current distortion correction and robust tensor estimation for MR diffusion imaging. In Niess, W., ed.: MICCAI'01. Springer, Berlin (1999)
6. Skare, S., Andersson, J.L.: Simultanous correction of eddy currents and motion in DTI using the residual error of the diffusion tensor: Comparison with mutual information. In: Proceedings of the ISMRM, Hawaii (2002) 508
7. Andersson, J.L., Skare, S.: A model-based method for retrospective correction of geometric distortions in diffusion-weighted EPI. NeuroImage **16** (2002) 177–199
8. Netsch, T., Rösch, P., van Muiswinkel, A., Weese, J.: Towards real-time multimodality 3-D medical image registration. In: Eighth International Conference on Computer Vision. Volume 1., Vancouver, BC (2001) 718–725
9. Holden, M., Hill, D.L., Denton, E.R., Jarosz, J.M., Cox, T.C., Rohlfing, T., Goodey, J., Hawkes, D.J.: Voxel similarity measures for 3-D serial MR brain image registration. IEEE Transactions on Medical Imaging **19** (2000) 94–102
10. Weese, J., Rösch, P., Netsch, T., Blaffert, T., Quist, M.: Gray-value based registration of CT and MR images by maximization of local correlation. In Taylor, C., Colchester, A., eds.: MICCAI'99. Springer, Berlin (1999) 656–663
11. Netsch, T., Weese, J., Rösch, P., van Muiswinkel, A., Desmedt, P.: Grey value-based 3-D registration of functional MRI time-series. In: Image Processing. Volume 3979 of SPIE Proceedings. (2000) 1148–1159
12. Studholme, C., Hill, D.L., Hawkes, D.J.: Automated 3-D registration of MR and CT images of the head. Medical Image Analysis **1** (1996) 163–175

Rigid Registration of Echoplanar and Conventional Magnetic Resonance Images by Minimizing the Kullback-Leibler Distance

Salil Soman[1,2,3], Albert C.S. Chung[4], W. Eric L. Grimson[1], and William M. Wells III[1,2,5]

[1] MIT Artificial Intelligence Laboratory, Cambridge, MA USA
[2] Surgical Planning Laboratory, Brigham & Women's Hospital, Boston, MA USA
[3] UMDNJ – Robert Wood Johnson Medical School, NJ USA
[4] Dept. of Computer Science, Hong Kong University of Science & Technology, HK
[5] Harvard Medical School, Boston, MA USA
sal@ai.mit.edu

Abstract. Functional Magnetic Resonance Imaging (fMRI) studies are derived from a time series of Echo-Planar images (EPIs). Compared to conventional Magnetic Resonance Images (MRIs), EPIs are of relatively poor quality for discerning anatomic features and are often registered with corresponding MRIs to map brain activity to neuroanatomy. In this paper we demonstrate the utility of a technique to register an EPI-MRI pair by minimizing the discrepancy between its joint intensity probability mass function (PMF) and a previously learned one for a properly registered EPI-MRI pair, using the Kullback-Leibler Distance (KLD). In probing experiments Joint Entropy (JE) and Mutual Information showed significant bias relative to KLD along the axial direction and JE along a rotation axis. A comparison of searches using random starting poses showed KLD to have lower final pose errors than JE. Results of variation on parameters of the KLD based EPI-MRI registration technique are also presented.

1 Introduction

Researchers often use functional Magnetic Resonance Imaging (fMRI) to study brain activity. These studies produce spatial activation maps through analysis of a time series of Echo-Planar Images (EPIs) taken of the patient's head as the subject responds to specific tasks or stimuli. While EPIs provide brain activity information on a time scale adequate for making inferences about regional brain function, they are of relatively poor quality for discerning anatomical features. Because there are a variety of tasks that are facilitated by the ability to correlate features like motor function to specific neuroanatomy (e.g. planning neurosurgical tasks or performing neuroscience experiments), researchers often desire to register EPIs to their corresponding Magnetic Resonance Images (MRI). This registration allows brain activity noted in regions of EPIs over successive volumes of the time series to be attributed to anatomic regions of the brain. Registration techniques which employ Mutual Information (MI) provide a partial solution to this problem, but they show limitations in capture range for rigid transformations. In this paper we demonstrate the utility of a method that leverages *a priori* domain knowledge to register an EPI –MRI pair by utilizing infor-

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 181–190, 2003.

mation about a properly registered EPI-MRI pair. The technique searches the space of transformations of the EPI of a test image pair for a pose that results in a joint intensity probability mass function (PMF) that most resembles a learned one for a properly registered training pair. The difference between the PMFs is determined using the Kullback-Leibler Distance (KLD), an information theoretic similarity measure commonly used in machine learning and information theory [1]. The KLD registration approach has previously been used to solve challenging angiographic registration problems [2, 3].

In fMRI studies a time series of EPIs of a patient's head are taken and regions of the EPIs are analyzed for changes over time [4]. Regions of successive EPIs where intensity changes correlate with the experimental protocol are indicative of brain activity. Brain activation indirectly causes intensity variations in the EPI images by way of the Blood Oxygenation Level Dependent (BOLD) effect [5]. Time series of EPIs must capture successive images at intervals of approximately once every two seconds to be able to reflect these rapid occurring changes in blood flow. Conventional MRIs are acquired over a period of some minutes per image capture, and provide more detailed images of brain anatomy [6]. Registering a set of EPIs to a set of corresponding MR images allows researchers to correlate areas of apparent activity in the EPI images with specific brain anatomy.

It is widely recognized that EPI images can contain significant distortions in addition to the intensity voids caused by magnetic susceptibility effects [6]. A definitive solution to the EPI/MRI registration problem will likely address this issue directly, by simulating the effect, or by field mapping at the time of acquisition. Nevertheless, we feel that in the interim, a robust method of rigid registration would be well received by the research community that uses fMRI.

One pragmatic approach to the MRI / EPI registration problem is to acquire a conventional MRI scan just before the EPI images for use as a registration reference (RR) that may be registered to other conventional MRI using, for example, MI. This method depends on the EPI and RR being in correspondence by the design of the scanning protocols. While the method is viable, it does require an additional scan, and it is not usable retrospectively for data that was acquired without the RR. In addition, there is frequently residual misregistration between the EPIs and the RR that needs to be corrected manually.

Many medical image registration problems have been solved using the Mutual Information approach [7, 8, 9]. The MI approach to registration seeks a transformation which maximizes the statistical dependence among the two images, without regard to particulars of the relationship. There are strong similarities between the MI approach and that of minimizing the joint entropy (JE). The objective functions share a joint entropy term, which is responsible for a pronounced extremum at the correct pose. The MI approach uses additional individual entropy terms that can enhance long range capture in some applications. In the experience of the authors, and as we will demonstrate below, the widely-used MI registration approach can, however, perform poorly on the EPI/MRI registration.

2 Methods

In this section we describe the objective function search and probing experiments reported in the results section.

2.1 Overview of Registration Experiments

In a training operation an EPI and its corresponding MRI are registered by an expert, blurred to 4 levels, and joint intensity PMFs are computed for all 5 of the EPI-MRI pairs. These PMFs will be referred to as Aligned Joint Intensity PMFs (AJPs).

In the registration phase, a test EPI-MRI pair is blurred to the same four levels that were used on the training pair. Then, starting with the most blurred image (blur level 4 in the figures), the algorithm searches for a rigid transformation of the test pair's EPI that results in a test pair joint intensity PMF that has a minimal KLD from the AJP for that blur level. Once this pose is found, the algorithm repeats this search on the test pair's EPI blurred to the highest blur level that has not yet been searched, using the final pose from the previous search as the start pose. The search continues until a pose is found to minimize the KLD between the test pair's PMF and the AJP of the non-blurred training pair. This final pose is returned as result.

2.1.1 Image Preparation and Selection

All image data in our experiments were acquired on a GE 1.5 Tesla Signa system equipped with the HORIZON hardware/software package. The MRI data in our experiments is a standard gradient echo (SPGR) scan that consists of 124 1.5 mm thick slices, FOV 24cm, image resolution 256 X 256 pixels. The EPI data sets contain 21 contiguous 7mm-slice images. The functional images are acquired in an auditory experiment using the EPIBOLD pulse sequence with the following set of parameters: TE=50msec, TR=3sec, FOV 24cm, image resolution=64 x 64 pixels.

EPI and MRI volumes were rigidly aligned for good visual agreement throughout, the EPI volume was re-sampled, using tri-linear interpolation, into the lattice of the MRI data, and corresponding 256 X 256 slice pairs were generated to be used as 2D test and training pairs.

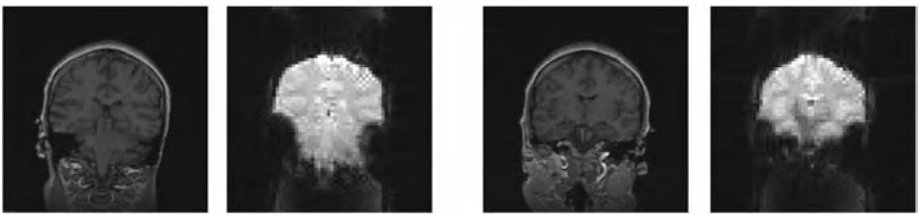


Fig. 1. MR and Echo-planar Image Pairs from A Single Acquisition Session

2.1.2 Blurring the Images

Four rotationally symmetric Gaussian low-pass filters of the following sizes and standard deviations (σ) were created to blur the original images: **Level 1**- 20x20 pixels, σ: 3, **Level 2** – 40x40 pixels, σ: 5, **Level 3** – 80x80 pixels, σ :8, **Level 4** – 100x100 pixels, σ: 10.

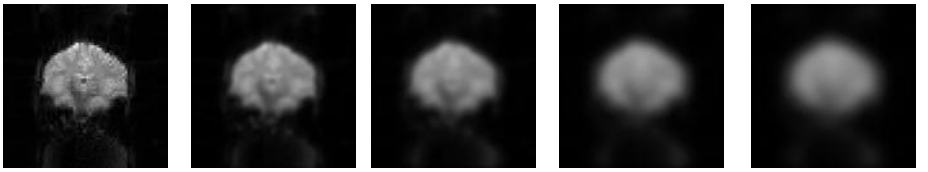


Fig. 2. (Left to Right) – Increasing strength blur applied to an Echo-planar image

2.1.3 Image Transformations

All transformations of the EPI images were performed using a rigid transformation model, with rotations occurring around the center of the image [10].

2.1.4 Histogramming to Estimate Joint Intensities

For the KLD objective function, a histogram was computed on an image pair by rastering over the EPI and MRI simultaneously and counting the co-occurrence of intensities in regions where the images overlapped. The co-occurring pixel intensity i from the EPI and j from the MRI led to incrementing bucket (i,j) in a 256 x 256 matrix.

The matrix was then normalized to convert the histogram into a PMF. Because it was possible for some buckets to have counts of zero after normalization, an ξ value was added to all buckets in the matrix before normalization. Varying ξ between $1x10^{-6}$ and $1x10^{-23}$ had no significant difference in the results of X probing experiments (see Objective Function Probing Experiments section). The experiments presented in this paper used an ξ of $1x10^{-23}$.

The histogramming techniques use for KLD and the JE or MI probes differed slightly. The JE and MI objective functions were computed in the space of the fixed image, with the moving image zero padded as necessary. Histograms for the MI and JE method were calculated with and without interpolation, respectively.

2.1.5 Kullback-Leibler Distance (KLD)

Given the learned $\hat{P}$ and the pose dependent observed P_o^T joint intensity PMFs, the Kullback-Leibler distance between the two PMFs is given by [11, 12]

$$D(P_o^T \parallel \hat{P}) = \sum_{i_1,i_2} P_o^T(i_1,i_2) \log \frac{P_o^T(i_1,i_2)}{\hat{P}(i_1,i_2)} \tag{1}$$

2.1.6 Interpolation vs. Rounding

Histograms were generated by applying a standard rounding function to image intensities. For comparison, we also implemented an interpolation function that spread counts of intensity co-occurrences over the buckets corresponding to the combinations of floors and ceilings of each intensity value.

2.1.7 Optimization of the Transformation T - Downhill Simplex Searches

The goal of the KLD based registration is to find the optimal transformation $\hat{T}$ by minimizing $\hat{T} = \arg\min_T D(P_o^T \| \hat{P})$ - the difference between the observed P_o and the expected $\hat{P}$ joint intensity PMFs.

For the registration experiments presented, the value of KLD or JE was minimized using the Nelder-Mead Downhill Simplex (direct search) method with a multi-resolution strategy [13, 14]. This approach does not use derivatives.

One hundred start poses were randomly chosen, where the X, Y, and θ parameters were varied within boundaries implied by Figure 3 to cause the majority, but not all downhill searches using KLD and JE to succeed. The parameters ranged from: -45 to 55 for X, -25 to 35 for Y, and -50º to 130º for θ. The KLD searches were then carried out as described in the overview part of the Methods section. The JE searches were performed in the same manner, except that the training phase was not carried out, and for each evaluation of the objective function, only the training image pair (transformed and / or blurred appropriately) was evaluated.

2.2 Objective Function Probing

Two EPI-MRI slice pairs were selected from a single acquisition set to be used as training and test pairs to perform rigid transformation probes (see Figure 1). Both image pairs were blurred to the four levels described earlier, and the AJPs for the blurred and non-blurred training pairs were computed. The probes then consisted of performing the following for the image pairs at all levels of blurring: transforming the test pair's EPI in fixed steps along a single axis (X, Y, or θ), computing the joint intensity PMF for the transformed EPI and its corresponding MRI blurred to the same level, and then computing the KLD between this PMF and the AJP for that blur level.

The X probe shifted the test pair EPI along the X axis from -65 to 65 pixels in 1 pixel increments, the Y probe performed the same shift along the Y axis, and the θ probe rotated the test EPI from -180 to180 degrees in two degree increments.

2.2.1 Intensity Scaling Image Pairs from Different Acquisitions

To account for inter-acquisition intensity variations for training and test pairs, experiments were performed in which an image dependent scaling factor was applied to all intensities during histogramming. The factor consisted of the number of buckets used divided by the intensity of the brightest pixel occurring in that image.

2.2.2 Pseudo 3D Experiment – Histogramming over Multiple Slices

For these experiments, X probes were performed as previously described, with the exception that every step requiring a single image pair used a set of images. Registered EPI-MRI pairs from a single time point were divided into two consecutive sets – anterior and posterior halves. Two probes were then performed, each in which one set was used for training and the other as a test pair. Generating the training AJPs entailed blurring all of the images in the training set, and then histogramming over all intensity pairs occurring in areas of image pair overlap. Computing the test pair PMF required transforming all EPIs of the test set and then histogramming cumulatively as described for generating the AJPs (See Figure 6).

3 Results

3.1 X Probe Comparisons: KLD, JE, MI

Joint Entropy [9], Mutual Information [15, 8] and KLD probes of X, Y or θ axes were performed on a test pair of images (Pair 12). MI & JE's definitions are given by

$$MI = \sum_{i_1,i_2} P_o^T(i_1,i_2) \log \frac{P_o^T(i_1,i_2)}{P_o^T(i_1)P_o^T(i_2)}, \text{ and } JE = -\sum_{i_1,i_2} P_o^T(i_1,i_2) \log P_o^T(i_1,i_2), \quad (2)$$

where $P_o^T(i_1)$ and $P_o^T(i_2)$ are the marginal distributions.

In Figure 3 the plots show the variations in the objective functions for each transformation of the EPI of the test pair for each of the blur levels. Negated MI is presented to facilitate comparison with JE and KLD, which both show minima near proper registration. The transformation resulting in the pose chosen by the expert is indicated by the vertical bar plus symbols. The square on each line indicates the pose resulting in the lowest value for the objective function. Joint Entropy showed significant bias on the Y and θ probes, while MI showed significant bias on the Y probe.

3.2 Objective Function Downhill Simplex Searches

Figure 4 summarizes the results of 100 Downhill simplex searches using KLD and JE, with the final pose error shown as a function of the start pose error. The start and final pose errors were defined as $\sqrt{Error_X^2 + Error_Y^2 + Error_\theta^2}$. These searches began with the test image pair blurred at the highest blur level and the EPI of the test pair transformed to the start pose. For a majority of the starting points, the KLD objective function returned a pose with a lower Final Pose Error than the JE based method. A final pose error threshold was selected to classify searches into successful and unsuccessful. This threshold selected the best-scoring significant groups of results. The average final pose for successful searches was then calculated. On average, successful KLD trials converged to pose parameters closer to those chosen by the expert (x=0, y=0, θ=0) than the JE trials, and showed smaller variation. JE was significantly biased in the Y axis, and showed its largest variation in the Y axis pose parameter. Figure 5 shows the overlay of the test EPI transformed to the pose chosen by the expert with its corresponding MRI (left image), the overlay for the test EPI transformed to the average pose parameters suggested by KLD (middle image) and by JE (right image). Visibly, the JE image is most discrepant from the expert's pose, with its registration aligning some brain activation to the scalp. The KLD based method showed a pose similar to the one chosen by the expert.

3.3 Effect of Different Training Pairs on KLD

We examined the effect of choice of training pair on KLD X probes. One EPI-MRI pair was designated as a test pair, and then two KLD X probes were performed, each

using a different EPI-MRI training pair (sampled from the same EPI-MRI volume pair). Both probes showed a minimum near the pose selected by the expert.

3.4 KLD Using Training and Test Pairs from Different Acquisitions

KLD X probes using a test pair from a similar but different EPI-MRI volume pair than the training pair, were performed directly and with the use of an image dependent scaling factor. Both probes yielded minima near the expert's pose for the non-blurred images, and showed minimal variations of minima for KLD at the four blur levels. The non-scaled probe showed more minima at the expert's pose, but the probe that scaled the image intensities showed less variation in the position of its minima. The final pose of either method would likely agree with the expert's.

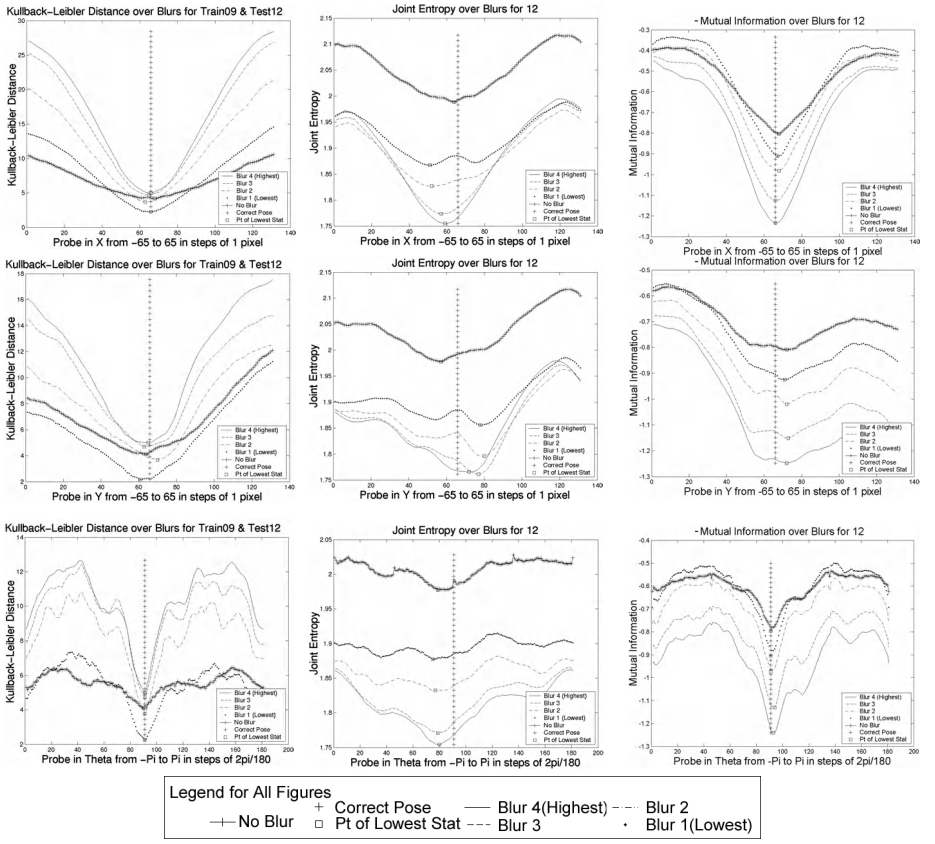


Fig. 3. Comparisons of X, Y, and θ Probes for registering an image pair using (left column) the Kullback-Leibler Distance (KLD), (middle column) Joint Entropy (JE), and (right column) Negated Mutual Information (MI). JE displayed significant bias in Y and θ, while MI shows significant bias along the Y axis

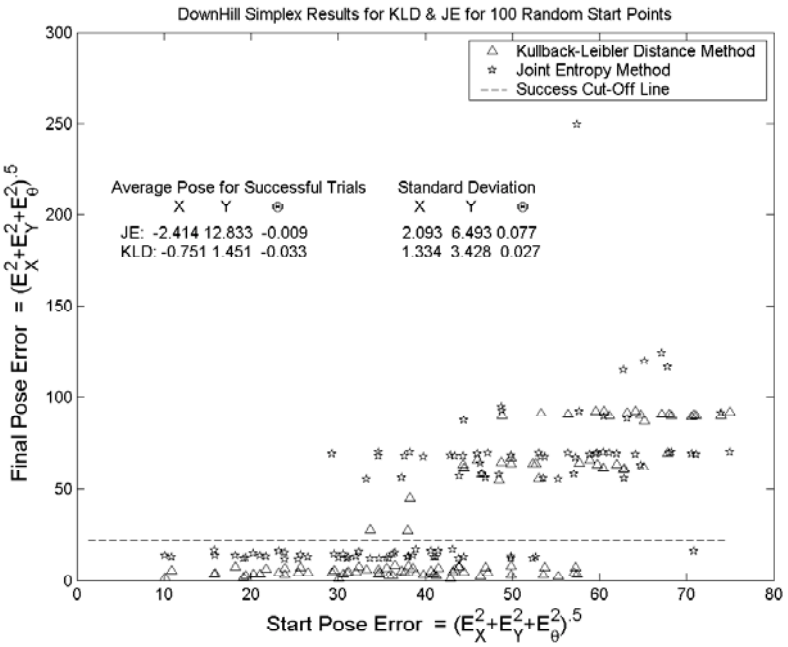


Fig. 4. Results of KLD and Joint Entropy Driven Downhill Simplex Searches for a test image given 100 random start poses for the test EPI. Trials were classified as being successful if the final pose error fell under a threshold. The KLD based successful searches showed average pose parameters closer to the expert's chosen pose than the Joint Entropy based method

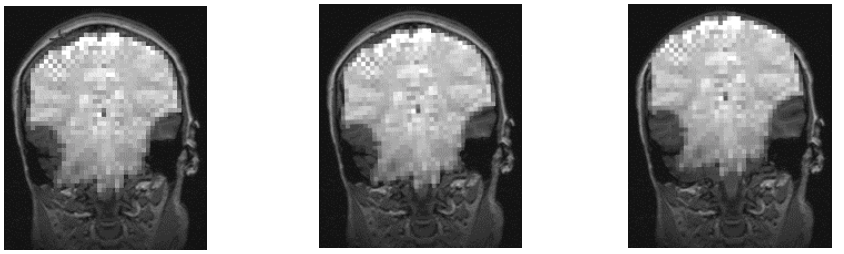


Fig. 5. Significant EPI intensities overlaid on corresponding MRI. (Left) Hand registration of the EPI-MRI pair by an expert, (middle) EPI transformed to the average final pose parameters of successful KLD method trials for downhill simplex, and (right) EPI transformed according to the average final pose parameters of successful Joint Entropy based method

3.5 Effects of Bucket Size and Interpolation on KLD

Non-integer intensity values were encountered during histogramming, and were rounded to allow bucketing. To see if rounding was leading to loss of important information, KLD X probes using a bi-linear interpolation of the intensity value were performed. The interpolation method did not show significant improvement of the

objective function. Similarly, increases in bucket size from 32 to 256 showed relatively small improvements in the smoothness or minima of the KLD X probes.

3.6 Combining Histograms across Multiple Slices – Pseudo 3D X Probes

Figure 6 shows the two KLD X pseudo 3D- probes performed. The left plot used the posterior half of the image set for training, and the anterior as the test set. The right plot used the opposite configuration. In general, both plots show much less variation in minima at all blur levels than the X probes performed using single training and test pairs. The right plot showed more uniform minima at the correct pose than the probe depicted on the left.

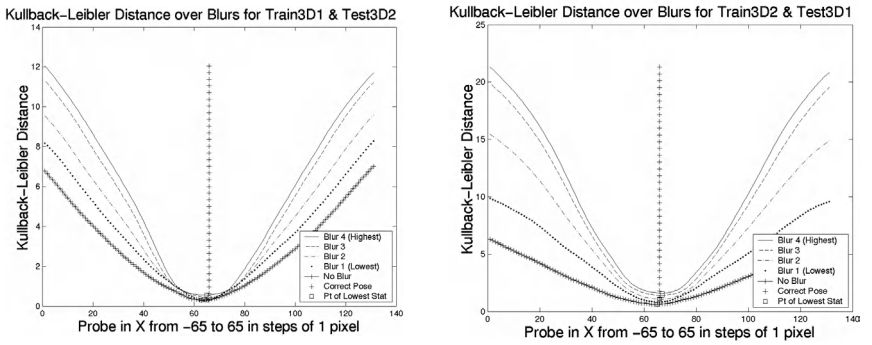


Fig. 6. Pseudo 3D Experiment Results – modified KLD X probes in which all the EPI-MRI pairs from one time point were divided into two sets (anterior and posterior halves), and information from one set was used for training, while the other was used as the test set. Intensity pair information from all the image pairs in a set was combined to make the histograms for these probes. Probes where the anterior half (right) and the posterior half (left) of the image set was used

4 Discussion

While MRI/EPI registration is a 3D-3D problem, in the present work, informative results have been obtained from a series of 2D experiments on representative slices. Clear shortcomings of the MI and JE objective functions were shown, especially in the axial direction, which are unlikely to be ameliorated by 3D registration. Variations on bucket size and the use of rounding or interpolation did not appear to significantly degrade performance of the KLD objective function. Similarly good results were achieved when using images from two different acquisition sets as training and test pairs to perform registration. The pseudo-3D experiments provide good indication that the KL approach will be useful for the full 3D-3D problem.

5 Future Work

This technique needs to be assessed within a three dimensional framework, where probing experiments and searches are performed over all six degrees of freedom.

Acknowledgements

Salil Soman is supported in part by a Medical Informatics training grant from the National Library of Medicine, and a grant from the Pfizer Corporation. We thank Cindy Wible for providing the image data. The following grants supported the work reported in this document - NLM, P41RR13218, P01CA67165, R21CA98449, and ERC9731748. AC would like to acknowledge support from RGC HKUST6209102E and SSRI01/02.EG22 grants.

References

1. Bishop, C. M. (1995). Neural networks for pattern recognition. Oxford, New York, Clarendon Press; Oxford University Press.
2. Albert C. S. Chung, William M. Wells III, Alexander Norbash, W. Eric L. Grimson: Multi-modal Image Registration by Minimizing Kullback-Leibler Distance. MICCAI (2) 2002: 525-532
3. Ho-Ming Chan, Albert C.S. Chung, Simon C.H. Yu, Alexander Norbash and William M. Wells III: Multi-modal image registration by Minimizing Kullback-Leibler distance between expected and observed joint class histograms. To appear in CVPR 2003
4. KK Kwong et al. Dynamic Magnetic Resonance Imaging of Human Brain Activity During Primary Sensory Stimulation. Proc. Natl. Acad. Sci USA, June 15, 1992.
5. S. Ogawa and T. M. Lee, "Magnetic resonance imaging of blood vessels at high fields: In vivo and in vitro measurements and image simulation," Magn. Reson. Med., vol. 16, pp. 9–18, 1990.
6. P. Jezzard and S. Clare. Sources of distortions in functional MRI data. *Human Brain Mapping*, 8:80-85, 1999.
7. Paul Viola , William M. Wells, III, Alignment by Maximization of Mutual Information, International Journal of Computer Vision, v.24 n.2, p.137-154, Sept. 1997
8. W.M. Wells, P. Viola, Viola P, Atsumi H, Nakajima S, Kikinis R. Multi-Modal Volume Registration by Maximization of Mutual Information. *Medical Image Analysis*, 1(1):35-51,1996.
9. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, P. Suetens, Multimodality image registration by maximization of mutual information , IEEE transactions on Medical Imaging, vol. 16, no. 2, pp. 187-198, April 1997
10. J. Maintz, and M. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1-36, 1998.
11. T.M. Cover and J.A. Thomas. *Elements of Information Theory.* John Wiley & Sons, Inc., 1991.
12. S. Kullback. *Information Theory and Statistics*. Dover Publications, Inc., 1968
13. Nelder, J. A., and Mead, R., A Simplex Method for Function Minimization, Computer Journal, Vol. 7, pp. 308-313, 1965.
14. W.H. Press, S.A. Teukolsky, et al. *Numerical Recipes in C, 2nd Edition.* Cambridge University Press, 1992.
15. Umaki, S. http://www.engineering.uiowa.edu/~aip/248_s02_solutions/soumik_ukil/hw3, May 2003

Non-rigid Spatio-Temporal Alignment of 4D Cardiac MR Images

Dimitrios Perperidis[1], Anil Rao[1], Raad Mohiaddin[2], and Daniel Rueckert[1]

[1] Visual Information Processing Group
Department of Computing
Imperial College London
180 Queen's Gate, London SW7 2BZ
United Kingdom
[2] Royal Brompton and Harefield NHS Trust
Sydney Street, London, United Kingdom

Abstract. In this paper we further develop of a 4D registration algorithm for the spatio-temporal alignment of cardiac MR image sequences.The registration algorithm has the ability not only to correct any spatial misalignment between the image sequences but also any temporal misalignment which maybe the result of differences in the cardiac cycle between subjects and differences in the temporal acquisition parameters. The algorithm uses a 4D transformation model which is separated into a spatial and a temporal component. In this approach the spatial part of transformation is composed by a global affine transformation and a local free-form deformation based on B-splines. The spatial part not only corrects spatial differences of a global nature but also local spatial differences of the cardiac shape. The temporal component uses an affine transformation which corrects the temporal misalignment caused by differences in the initial acquisition offset and length of the two cardiac cycles. The method was applied to a number of cardiac MR image sequences from healthy volunteers. The registration was qualitatively evaluated by visual inspection and quantitatively by measuring overlap of anatomical regions between the sequences. The results are compared with the results of the previously developed 4D registration method. A significant improvement in the alignment of the images is achieved by the use of free-form deformation models.

1 Introduction

Cardiovascular diseases are the leading cause of death in the UK [1]. Their early diagnosis and treatment is crucial in order to reduce mortality and to improve patients' quality of life. Recent advances in the development of non-invasive imaging modalities are enabling the high resolution imaging of the cardiovascular system. Among these modalities, magnetic imaging (MR) is playing an increasingly important role. MR imaging allows not only the acquisition of 3D images which describe the cardiac anatomy but also the acquisition of 4D cardiac image sequences which describe the cardiac anatomy as well as function.

These recent advances in the development of cardiac imaging modalities have led to an increased need for cardiac image registration methods (see [2], as well as the paper by Mäkelä et al. [3] for a recent review of cardiac image registration methods and [4,5,6] for

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 191–200, 2003.

a general review of image registration methods). In general, cardiac image registration is a very complex problem due to the complicated non-rigid motion of the heart and the thorax as well as the low resolution in which the cardiac images are usually acquired. In recent years cardiac image registration has emerged as an important tool for cardiac or respiratory motion correction [7,8] and cardiac motion estimation [9],[10]. Cardiac image registration also plays an important role for the fusion of information from a number of different image modalities such as CT, MR, PET and SPECT [11,12]. In addition, cardiac image registration is crucial for the comparison of images of the same image modality, e.g. before pharmacological treatment or surgical intervention and after the treatment. Other applications include the inter-subject registration of cardiac image sequences to the same coordinate space (anatomical reference) in order to enable a direct comparison between the cardiac anatomy and function of different subject [2,13,14]. Such inter-subject registration is the essential for the development of an anatomical and functional atlas of the cardiovascular system similar to atlases that have already been developed for the human brain [15].

In this paper we extend a 4D cardiac MR image registration method based on voxel-based similarity measures which has been recently presented [2]. This method will not only bring a number of sequences of cardiac images acquired from different subjects or from the same subject (for example short and long axis cardiac image sequences) into the same spatial coordinate frame but also into the same temporal coordinate system. The method not only allows direct comparison between the cardiac anatomy of different subjects to be made but will also allow comparisons between the cardiac function to be made. The aim of this contribution is to improve the accuracy of the cardiac MR image sequence registration algorithm by using a free-form deformation model for the spatial part of the transformation. Hence, the registration method has the ability to correct any spatial misalignment caused by both global and local difference between the shape of the hearts as well as any temporal misalignment due to differences in the temporal acquisition parameters. The remainder of the paper is organised as follows: section 2 provides a brief review of related work while section 3 describes the proposed 4D image registration algorithm. In Section 4 we present the results of the experiments performed in order to demonstrate the applicability of our method. Finally, Section 5 describes the conclusions from this work and describes the intended future work

2 Related Work

While a large number of registration techniques exists for cardiac images, most of these techniques focus on 3D images and ignore any temporal misalignment between the image sequences. In an earlier publication, [2], we developed a novel approach for the spatio-temporal registration of 4D cardiac MR images. The approach uses a 4D transformation model $\mathbf{T} : (x, y, z, t) \rightarrow (x', y', z', t')$ which maps any point of one image sequence $I(x, y, z, t)$ into its corresponding point in the reference image sequence $I(x', y', z', t')$. The mapping has been resolved into decoupled spatial and temporal components. The temporal component was represented as an affine transformation which corrected for temporal differences of affine nature while the spatial part of the transformation was also an affine transformation with 12 degrees of freedom utilising scaling, shearing, rotation

and translation [2]. The method was evaluated using cardiac MR images sequences from seven healthy volunteers. The results of the registration showed significant improvement in the spatio-temporal registration of the cardiac MR image sequences.

Klein et al. [16] investigated a 12 parameter global affine motion model for the registration of different respiratory gates in an end diastolic cardiac PET sequence. In order to obtain robust motion estimates, a 4D registration model was produced which encourages smoothly varying motion between adjacent frames. Similarly to their earlier work [17],[18], this approach assumes that the motion of the heart from one frame to the other is likely to follow a uniform motion. The approach uses priori knowledge of the cardiac motion during the registration of two respiratory gated images and as a result there is significant reduction of noise and blurring in the registered image.

Outside cardiac image analysis, Caspi and Irani [19] developed an approach for the spatio-temporal alignment of image sequences. Their paper studies the problem of establishing correspondence in both the temporal and the spatial domains between two video sequences of the same scene. They solve the alignment problem by dividing it into two components: the spatial misalignment which is caused by the external and internal calibration parameters and the temporal misalignment which is caused when the two image sequences have a time shift or different frame rates. Hence, the transformation they used has two parts: a temporal and a spatial part. The spatial part of the transformation is a 2D affine and the temporal part of the transformation is an 1D affine transformation in time [19]. They have developed two different sequence to sequence alignment methods: the first method aligns the two sequences by aligning the trajectories of manually marked points while the other one uses the only grey level information of the two sequences.

3 Spatio-Temporal Registration of 4D Cardiac Image Sequences

Since the heart is undergoing a spatially and temporally varying degree motion during the cardiac cycle, 4D cardiac image registration algorithms are required when registering two cardiac MR image sequences. Spatial alignment of corresponding frames of the image sequences (e.g. the second frame of one image sequence with the second frame of the other) is not enough since these frames may not correspond to the same position in the cardiac cycle of the image sequences. Spatio-temporal alignment of cardiac MR image sequences will resolve spatial ambiguities which occur when there is not sufficient common appearance in the two 3D MR cardiac images. Furthermore, it can also improve the results of the registration because it is not restricted only to the alignment of existing frames but it can also use sub-frame information.

A 4D cardiac image sequence can be represented as sequence of n 3D images $I_k(x, y, z)$ with a fixed field of view Ω_I and an acquisition time $t_k, t_k < t_{k+1}$, in the temporal direction. The resulting image sequence can be viewed as 4D image $I(x, y, z, t)$ defined on the spatio-temporal domain $\Omega_I \times [t_1, t_n]$. The goal of 4D image registration described in this paper is to relate each point of one image sequence to its corresponding point of the reference image sequence. In this case the transformation $\mathbf{T} : (x, y, z, t) \rightarrow (x', y', z', t')$ maps any point of one image sequence $I(x, y, z, t)$ into its corresponding point in the reference image sequence $I(x', y', z', t')$. The 4D mapping used in this paper is of the form:

$$\mathbf{T}(x, y, z, t) = (x'(x, y, z), y'(x, y, z), z'(x, y, z), t'(t)) \tag{1}$$

and can be of a subvoxel displacement in the spatial domain and of a sub-frame displacement in the temporal domain. The 4D mapping can be resolved into decoupled spatial and temporal components $\mathbf{T}_{spatial}$ and $\mathbf{T}_{temporal}$ respectively where

$$\mathbf{T}_{spatial}(x, y, z) = (x'(x, y, z), y'(x, y, z), z'(x, y, z))$$

and

$$\mathbf{T}_{temporal}(t) = t'(t)$$

each of which we choose to be one-to-one mappings. One consequence of this decoupling is that each temporal frame t in image sequence I will map to another temporal frame t' in image sequence I', ensuring causality and preventing different regions in a 3D image $I_t(x, y, z)$ from being warped differently in the temporal direction by $T_{temporal}$.

3.1 Temporal Alignment of 4D Image Sequences

The alignment of the two 4D cardiac MR images involves a temporal alignment and a spatial alignment. The temporal alignment of two image sequences aims to find the transformation function $\mathbf{T}_{temporal}$ which establishes correspondences between time t in one image sequence and the time t' in the reference image sequence, thereby it provides a mapping between corresponding time points in two cardiac cycles. In addition to differences in the length of the cardiac cycle, the temporal alignment of two image sequences is further complicated by the fact that the acquisition of cardiac MR images typically depends on two parameters: the first parameter describes the delay t_{delay} after the R-wave after which the MR acquisition starts while the second parameter Δt describes the temporal resolution of the image sequence. Figure 1 demonstrates how a typical ECG-gates cardiac MR image sequence is acquired. Together with the differences in the length of the cardiac cycle between subjects and differences in the acquisition parameters account for the temporal misregistration of the image sequences.

In our current implementation the temporal transformation, $\mathbf{T}_{temporal} : (t) \rightarrow (t')$, is represented as an affine transformation of the following form:

$$\mathbf{T}_{temporal}(t') = \alpha t + \beta \tag{2}$$

Here α accounts for scaling differences between the two image sequences while β accounts for differences in the translation between the two image sequences.

3.2 Spatial Alignment of 4D Image Sequences

The aim of the spatial part of the transformation is to relate each spatial point of an image to a point of the reference image, i.e. $\mathbf{T}_{spatial} : (x, y, z) \rightarrow (x', y', z')$ maps any point (x, y, z) of a particular time frame t in one image sequence into its corresponding point (x', y', z') of another particular time frame t' of the reference image sequence. The transformation $\mathbf{T}_{spatial}$ consists of a global transformation and a local transformation:

$$\mathbf{T}_{spatial}(x, y, z) = \mathbf{T}^{global}_{spatial}(x, y, z) + \mathbf{T}^{local}_{spatial}(x, y, z) \tag{3}$$

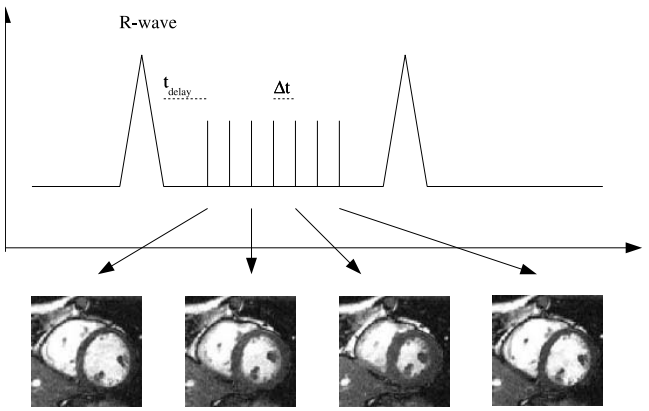


Fig. 1. An example of an ECG-gated acquisition of a cardiac MR image sequence: The parameter t_{delay} describes the delay after the R-wave after which the MR acquisition starts while Δt describes the temporal resolution of the image sequence.

The global transformation addresses differences in the size, orientation and alignment of the hearts while the local part addresses differences in the shape of the hearts. The simplest choice of $\mathbf{T}_{spatial}^{global}$ is a rigid transformation which has six degrees of freedom corresponding to translation and rotation. We choose an affine transformation which is a more general class of transformations with 12 degrees of freedom utilising scaling and shearing in addition to translation and rotation. Such a 3D affine transformation can be written as:

$$\mathbf{T}_{spatial}^{global}(x,y,z) = \begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \\ \theta_{31} & \theta_{32} & \theta_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} \theta_{14} \\ \theta_{24} \\ \theta_{34} \end{pmatrix} \tag{4}$$

where the coefficients θ parameterise the twelve degrees of freedom of the transformation.

The affine transformation describes the global spatial differences between two subjects. A free form deformation (FFD) model based on B-splines is used in order to describe difference in the local shape of the hearts. This kind of model has been successfully used in a number of non-rigid registration approaches [20,14,9]. To define a spline-based FFD we denote the spatial domain of the image volume as $\Omega_I = \{(x,y,z) \mid 0 \leq x < X, 0 \leq y < Y, 0 \leq z < Z\}$. Let Φ denote a $n_x \times n_y \times n_z$ mesh of control points $\phi_{i,j,k}$ with uniform spacing δ. Then, the FFD can be written as the 3D tensor product of the familiar 1D cubic B-splines [20]:

$$\mathbf{T}_{spatial}^{local}(x,y,z) = \sum_{l=0}^{3}\sum_{m=0}^{3}\sum_{n=0}^{3} B_l(u)B_m(v)B_n(w)\phi_{i+l,j+m,k+n} \tag{5}$$

where $i = \lfloor \frac{x}{n_x} \rfloor - 1, j = \lfloor \frac{y}{n_y} \rfloor - 1, k = \lfloor \frac{z}{n_z} \rfloor - 1, u = \frac{x}{n_x} - \lfloor \frac{x}{n_x} \rfloor, v = \frac{y}{n_y} - \lfloor \frac{y}{n_y} \rfloor, w = \frac{z}{n_z} - \lfloor \frac{z}{n_z} \rfloor$ and where B_l represents the l-th basis function of the B-spline:

$$B_0(u) = (1-u)^3/6$$
$$B_1(u) = (3u^3 - 6u^2 + 4)/6$$
$$B_2(u) = (-3u^3 + 3u^2 + 3u + 1)/6$$
$$B_3(u) = u^3/6$$

B-splines are locally controlled which makes them computationally efficient even for a large number of control points. In particular, the basis functions of cubic B-splines have a limited support, i.e. changing control point $\phi_{i,j,k}$ affects the transformation only in the local neighbourhood of that control point.

3.3 Similarity Measures for 4D Image Sequences

The optimal transformation is found by maximising a voxel based similarity measure. The use of a voxel-based similarity measure eliminates the need for any feature detection and segmentation of structures such as the epi- or endocardial surfaces. We choose normalised mutual information (NMI) [21] as a measure of spatio-temporal alignment which has been previously used successfully for mono- and multi-modality image registration. The normalised mutual information of two image sequences can be written as

$$I(A,B) = \frac{H(A) + H(B)}{H(A,B)} \tag{6}$$

where $H(\cdot)$ denotes the marginal entropy of an images sequence and $H(\cdot,\cdot)$ is the joint entropy of two image sequences. The normalised mutual information of the two image sequences can be calculated directly from the joint intensity histogram of the two sequences over the spatio-temporal domain of overlap $\Omega_{I_A} \times [t_{A1}, t_{An}] \bigcap \mathbf{T}(\Omega_{I_B} \times [t_{B1}, t_{Bn}])$. During the optimisation new voxel values are generated in the temporal domain using linear interpolation and trilinear interpolation in the spatial domain. The optimisation is carried out using an iterative downhill descent algorithm to calculate the optimal transformation:

$$\arg \max_{\alpha,\beta,\Theta,\Phi} I(A, T(B)) \tag{7}$$

where A is the reference image sequence and $\mathbf{T}(B)$ is the transformed image sequence B. The first step of the optimisation procedure optimises NMI as a function of the global spatial and temporal transformation, i.e. we optimise α, β, Θ in equation 7. The second step of the optimisation procedure optimises NMI as a function of the local spatial transformation, i.e. Φ in equation 7. A simple iterative gradient descent method which steps in the direction of the gradient vector, $\nabla \mathcal{C} = \frac{\partial \mathcal{C}(\Theta, \Phi^l)}{\partial \Phi^l}$, with a step size μ was used as optimisation procedure.

4 Results and Discussion

To evaluate the non-rigid spatio-temporal registration algorithm we have acquired cardiac MR image sequences from seven volunteers. All image sequences used for our experiments were acquired on a Siemens Sonata 1.5 T scanner using TrueFisp pulse sequence. For the reference subject 32 different time frames were acquired (cardiac cycle

of length 950msec). Each 3D image of the sequence had a resolution of $256 \times 192 \times 46$ with a pixel size of $0.97mm \times 0.97mm$ and a slice thickness of 3mm. Six 4D cardiac MR images were registered to the reference subject. The length of the cardiac cycle of these six images varied from 300msec to 800msec. As mentioned above, we have first carried out a spatial and temporal affine registration followed by an optimisation of the spatial local free-form transformation parameters.

The figure 2 provides an example of the registration. In the first row of figure 2 (a,b,c) are the short axis (a) and the long axis (b) views of particular slices of the reference sequence and also a temporal view of a vertical short axis line (c) of the reference sequence. The middle and bottom row contains the corresponding views of the corresponding slices after the affine registration [2] and after the non-rigid registration respectively. In the images of the middle and bottom row the anatomy contours of the reference subject are also overlaied. The figure shows a large improvement in the alignment of the images after the non-rigid 4D cardiac image registration. In the images (c,f,i) of figure 2 we can see how well the temporal alignment of the two sequences has been achieved.

The quality of the registration was measured by calculating the volume overlap for the left and right ventricle as well as for the myocardium (defined by an expert observer), $min(\Delta(T,S), \Delta(S,T)$. The volume overlap for an object O is defined as:

$$\Delta(T,S) = \frac{|T \bigcap S|}{|T|} \times 100\% \tag{8}$$

Here T denotes the voxels in the reference (target) image part of object O and S denotes the number of voxels in the other image part of object O. Table 1 shows the mean volume overlap and the standard deviation for each anatomical region after the non-rigid registration and after affine registration ([2]). The results indicate clearly that the use of the non-rigid spatial part provides a significant improvement in the quality of the registration.

Table 1. The mean volume overlap before the affine 4D registration and after non-rigid 4D registration.

Anatomical region	**Non rigid 4D registration (SD)**	**Affine 4D registration (SD)**
Left ventricle	83.00% (3.83%)	75.34% (7.09%)
Right ventricle	70.16% (4.50%)	65.23% (6.92%)
Myocardium	65.80% (4.53%)	61.32% (5.05%)

It is not expected that the images would be perfectly aligned in the temporal domain due to the nature of the temporal transformation, thus correcting differences in the trigger delay between the two sequences as well as differences in the length of the cardiac cycle. Therefore, it would not be able to address the temporal misalignment caused by differences in the dynamic properties of the contraction and relaxation phases of the cardiac cycle (one heart may have longer contraction phase than relaxation phase). An example of such a case is provided in [2]. However, the use of the non-rigid spatial

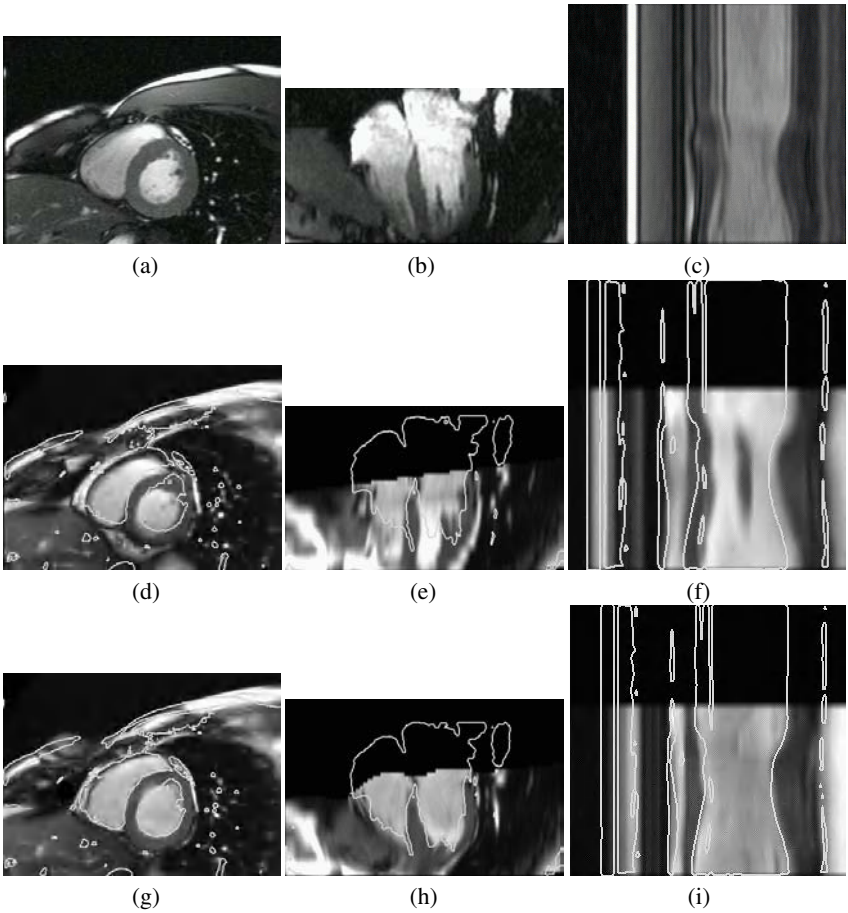


Fig. 2. Results of the 4D cardiac MR registration algorithm (a) shows the short axis view of the reference subject, (b) shows the long axis view of the reference subject (c) a temporal view of a short axis line of the reference subject (d-e) shows the corresponding short axis and long axis views after the affine registration (f) shows the corresponding temporal view of a short axis line after the affine registration (g)-(i) show the corresponding images after the 4D non-rigid registration.

transformation has enabled us to significantly improve the registration between the image sequences by taking into account the non-rigid differences between the subjects.

5 Conclusions and Further Work

A 4D transformation model which consists of decoupled spatial and temporal components has been proposed in an iterative registration approach for the non-rigid spatio-temporal alignment of two cardiac MR image sequences. The proposed registration approach uses normilised mutual information as a measure of image sequence alignment and enables the comparisons of the anatomy and the cardiac function of a number

of MR cardiac image sequences. The transformation model corrects both the temporal misalignment (caused by differences in the initial triger delay and in the length of the two cardiac cycles) and the spatial misalignment (caused by spatial differences of global nature and local nature between the two image sequences). The quality of the spatio-temporal aligiment of the image sequences has been significantly improved by the use of a spatial part which consist of an affine global transformation and a local free form deformation. The method shows a significant improvement on the registration over the previously presented 4D cardiac MR image registration algorithm [2]. The spatio-temporal alignment can be further improved by employing a more complex transformation in which both the spatial and the temporal component will be modelled by B-spline free-form deformation models. This registration method can be used for the inter-subject spatio-temporal alignment of a number of cardiac MR image sequence in order to construct a probabilistic atlas of cardiovascular system.

References

1. American Heart Association. Heart and stroke statistical update. `http://www.americanheart.org/`, 2002.
2. D. Perperidis, A. Rao, M. Lorenzo-Valdés, R. Mohiaddin, and D. Rueckert. Spatio-temporal alignment of 4D cardiac MR images. In *Second International Workshop on Functional Imaging and Modelling of the Heart, FIMH'03*, Lyon, France, June 5-6, 2003.
3. T. Mäkelä, P. Clarysse, N. Sipila, O.and Pauna, and Q. C. Pham. A review of cardiac image registration methods. *IEEE Transcactions on Medical Imaging*, 21(9), 2002.
4. J.B.A Maintz and M.A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–37, 1998.
5. D.L.G. Hill, P.G. Batchelor, M. Holden, and D.J. Hawkes. Medical image registration. *Physics in Medical Biology*, 46(3):R1–R45, 2001.
6. H. Lester and S. Arridge. A survery of hierarchical non-linear medical image registration. *Pattern Recognition*, 32(1):129–149, 1999.
7. D. Manke, P. Rosch, K. Nehrke, P. Bornert, and O. Dossel. Model evaluation and calibration for prospective respiratory motion correction in coronary MR angiography based on 3D image registration. *IEEE Transcactions on Medical Imaging*, 21(9), 2002.
8. K. McLeish, D.L.G Hill, D. Atkinson, J.M. Blackall, and R. Razavi. A study of the motion and deformation of the heart due to respiration. *IEEE Transcactions on Medical Imaging*, 21(9), 2002.
9. R. Chandrashekara, R. H. Mohiaddin, and D. Rueckert. Analysis of myocardial motion in tagged MR images using nonrigid image registration. In M. Sonka and J. Michael Fitzpatrick, editors, *Proceedings of the SPIE International Symposium on Medical Imaging*, pages 1168–1179, San Diego, California USA, 24–28 February 2002. SPIE.
10. R. Chadrashekar, A. Rao, G. Sanchez-Ortiz, R. Mohiaddin, and D. Rueckert. Construction of a statistical model for cardiac motion analysis using non-rigid image registration. In *Information Processing in Medical Imaging 2003*, Amleside, UK, 20-25 July 2003.
11. M.C. Gilardi, G. Rizzo, A. Savi, C. Landoni, V. Bettinardi, C. Rosseti, G. Striano, and F. Fazio. Correlation of SPECT and PET cardiac images by a surface matching registration technique. *Computerized Medical Imaging and Graphics*, 22(5):391–398, 1998.
12. T. G. Turkington, T. R. DeGradom, M. W. Hanson, and E. R. Coleman. Alignment of dynamic cardiac PET images for correction of motion. *IEEE Transaction on Nuclear Science*, 44(2):235–242, 1997.

13. L.M. Bidaut and J.P. Vallëe. Automated registration of dynamic MR images for the quantification of myocardial perfusion. *Journal Of Magnetic Resonance Imaging*, 13(4):648–655, 2001.
14. A. Rao, G.I. Sanchez-Ortiz, R. Chandrashekara, M. Lorenzo-Valdés, R. Mohiaddin, and D. Rueckert. Comparison of cardiac motion across subjects using non-rigid registration. In *Proceedings: Medical Image Computing Assisted Intervention, MICCAI 2002*, Tokyo, Japan, 2002.
15. J.C Mazziotta, A.W. Toga, A. Evans, P. Fox, and J. Lancaster. A probabilistic atlas of the human brain:theory and rationale for its development. The international consortium for brain mapping. *Neuroimage*, 2:89–101, 1995.
16. G.J. Klein, B.W. Reutter, and R.H. Huesman. Four dimensional affine registration models for respiratory-gated PET. *IEEE Transactions on Nuclear Science*, 48(3):756–760, 2001.
17. G.J. Klein. Four dimensional processing of deformable cardiac PET data. In *workshop on Mathematical Methods in Biomedical Image Analysis June 11-12 2000*, pages 127–134, Hilton Head Island, SC, 2000.
18. G.J. Klein and R.H Huesman. Four dimensional processing of deformable cardiac PET data. *Medical Image Analysis*, 6(1):29–46, 2002.
19. Y. Caspi and M. Irani. Spatio-temporal alignment of sequences. *IEEE Transactions on Pattend Analysis and Machine Intelligence*, 24(11):1409–1424, 2002.
20. D. Rueckert, L. I. Sonoda, C. Hayes, D.L.G Hill, M.O. Leach, and D.J Hawkes. Non-rigid registration using free-form deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.
21. C. Studholme, D.L.G. Hill, and D.J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32:71–86, 1999.

Digital Anatomy Atlas and Its Registration to MRI, fMRI, PET: The Past Presents a Future

Ruzena Bajcsy

CITRIS (Center for Information Technology Research in the Interest of Society)
University of California, Berkeley, CA 94720-1764.

This paper is dedicated with deep appreciation to all my former students, collaborators and researchers who use information technology to make the lives of others better.

1 Introduction

The purpose of this paper is to review the past and present work in Digital Anatomy of the Human Brain and the registration methods developed for its use, then to extrapolate where the new open problems are, as well as the opportunities in this field. Of course, it will be a very personal review with the focus of my work and of my collaborators with reference to the many others who built upon this early effort.

Naturally we will not be able to do justice to all relevant works in this area since the field is flourishing beyond our wildest expectations. However, in spite of all the work that has taken place, there are still many challenges left, especially as new imaging devices are coming into use, we need to look at how to register different scales and modalities.

2 The Past

We began this effort around 1978-1979 via the Ph.D. work of Chaim Broit [6]. We have been exposed to the problem of how to recognize/delineate regions of interest (ROI) in Computer Tomographic (CT) images and Positron Emission Tomographic (PET) images of the human brain as part of the discussion of the group at the Cerebrovascular Research Center in the Medical School at the University of Pennsylvania in Philadelphia around 1977, led by professors Reivich and Chance. Listening to the discussion of the researchers, it became very evident that the doctors referred to different ROIs in terms of anatomic structures, which every medical student learns as part of their training. Hence, it became clear that any systematic delineation of anatomic structures must be guided by an anatomic atlas of the human brain.

Furthermore, any recognition of the ROI in the PET images, which represent physiology rather than anatomy, had to be explained in terms of the corresponding anatomy. Toward the goal of creating a digital anatomy atlas, we used stained slices of the brain of a dead soldier from the so-called Yakovlev Collection of the Armed Forces Institute of Pathology [7]. This normal 31 year old male brain was embedded in celloidin and cut parallel to the plane bisecting the anterior and posterior commis-

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 201–211, 2003.

sures. Sections were 35 μm thick and every 20th section was stained for myelin. (Continuous sections were stained with cresyl violet, so the potential exists to go back to them in order to define cyto-architectural regions) This resulted in tissue sections that were 700 μm apart, which corresponded to 0.89 mm in vivo after correcting for the shrinkage due to the embedding and staining processes. Each section was photographed from a fixed location, with a ruler included for later scaling. The negatives were enlarged onto dimensionally stable acetate sheets and these were then digitized using a video camera connected to a Vax computer. The images of the rulers were carefully measured and any slight differences in any of the photographic steps were corrected by appropriate software scaling.

After the tedious digitization process (only the outlines of structures were digitized) of slices (see Figures 1, 2, 3), we reconstructed the three-dimensional atlas by stacking registered slices (see Figures 6, 7, 8). The outlines were not sufficient, so we needed to encode the tissue color corresponding to gray values of the CT images, as shown in Figures 4 and 5.

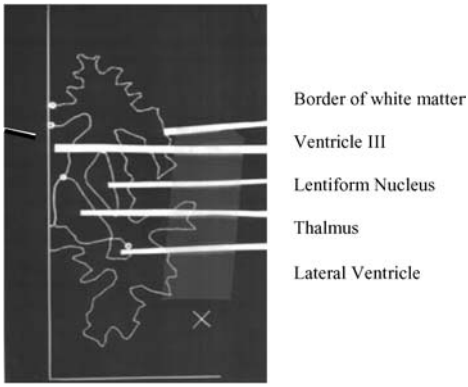


Fig. 1. Example of digitization process

2.1 The Registration Process

The standard approach of the time was to analyze images slice by slice. In fact, this approach was enhanced by the fact that the resolution between the slices was much worse (almost 10 x) than the x,y resolution. However, after a short examination of the problem, it was clear to us that due to different head position in the scanner, the slices were distorted. Hence, Broit [6] and I proposed that registration take place in volume rather than in two dimensions (see Figure 9 from Three-Dimensional Analysis and Display of Medical Images, R. Bajcsy, in Positron Emission Tomography, 119-129;1985).

Additional distortion was recognized due to individual anatomic differences. This lead us to introduce an elastic matcher, that is, the atlas was modeled as a rubber sheet. The underlined assumption of our registration process was that there exists enough invariant features in each individual brain that can be identified in the atlas and they will serve as fixed points for registration. The remaining structures in between the fixed points will be adjusted elastically following the Rubbersheet equations [1].

$$C_1 \nabla^2 U + C_2 \frac{\partial O}{\partial X} + F_x = 0$$

$$C_1 \nabla^2 V + C_2 \frac{\partial O}{\partial Y} + F_y = 0$$

$$C_1 \nabla^2 W + C_2 \frac{\partial O}{\partial Z} + Fz = 0$$

Rubbersheet equations (1)

Where X, Y, Z is a point in one image before deformation, U, V, W is the same point's coordinates following deformation, C_1 and C_2 are the elastic constants of the material and θ is the volume explanation at (X,Y,Z) and F_x, F_y, and F_z are the external vector force at each point.

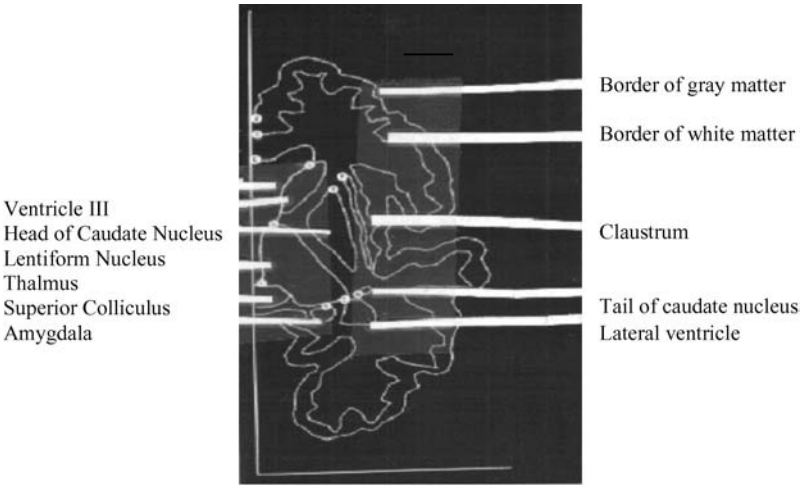


Fig. 2. Completed horizontal section of a human brain at OM4

The registration process consisted of 3 steps:

Step1: Global matching. The assumption was that both the atlas and the CT volume can be approximated by an ellipsoid and, hence we performed translation, rotation and scale transformation on the atlas for the best match to the data.

Step 2: Divide both the atlas and the CT data into cube windows and in each window compute a feature vector corresponding to edges in 6 different orientations (see Figure 10 from Three-Dimensional Analysis and Display of Medical Images, R. Bajcsy, in Positron Emission Tomography, 119-129;1985).

Step 3: Find the best local correspondence and perform the elastic deformation on the atlas data such that it minimizes the difference between the CT volume and the atlas volume. The results of the match are shown in figures 11 and 12.

Fig. 3. Reflected coronal section from a cat brain, where the head of the Caudate nucleus is highlighted

The next question was validation of these results. In the paper [7], we recruited five professionals (one neurologist, an anatomist, a neuroscientist and two radiologists) and had them outline the anatomic structures and compare their findings with what the machine found. The results from the machine match were comparable to the professional's match, including mistakes.

The difficulties we had were due to coarse resolution of the images (as obvious from the pictures above) and to the small amount of memory available at that time at PDP-11/60 computer (32K words). The projection of registered structures onto PET images are seen in Figure 13.

Fig. 4. Reflected horizontal section from Figure 2

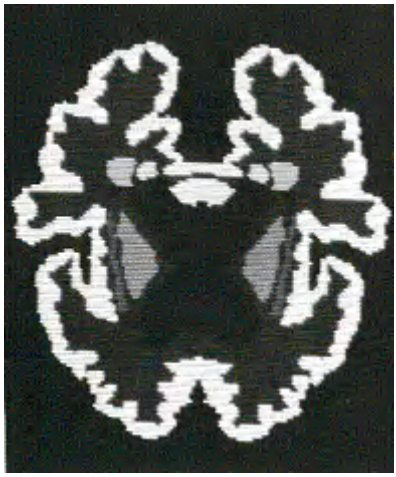


Fig. 5. Point-by-point representation of Figure 4. Different gray shadings define the areas included within individual anatomical structures

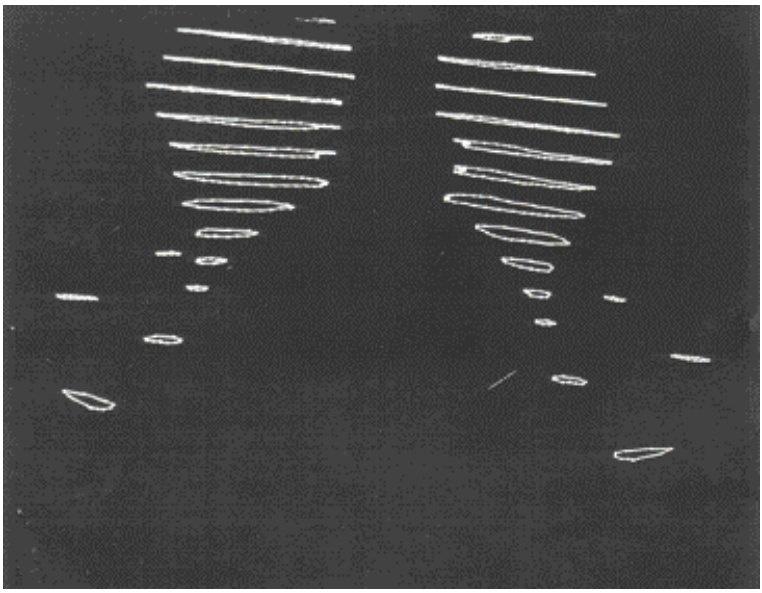


Fig. 6. Frontal view of stacked caudete nucleus through 13 slices

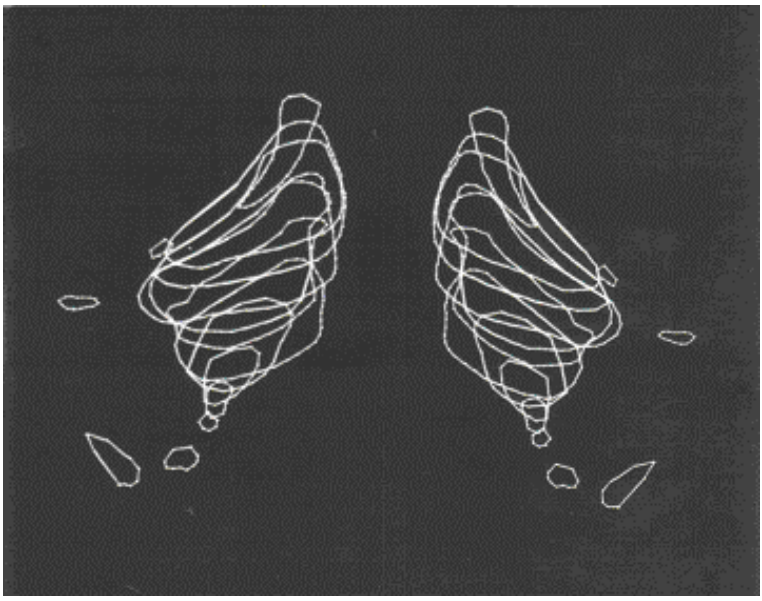


Fig. 7. Rotated view of about 90º of Figure 6

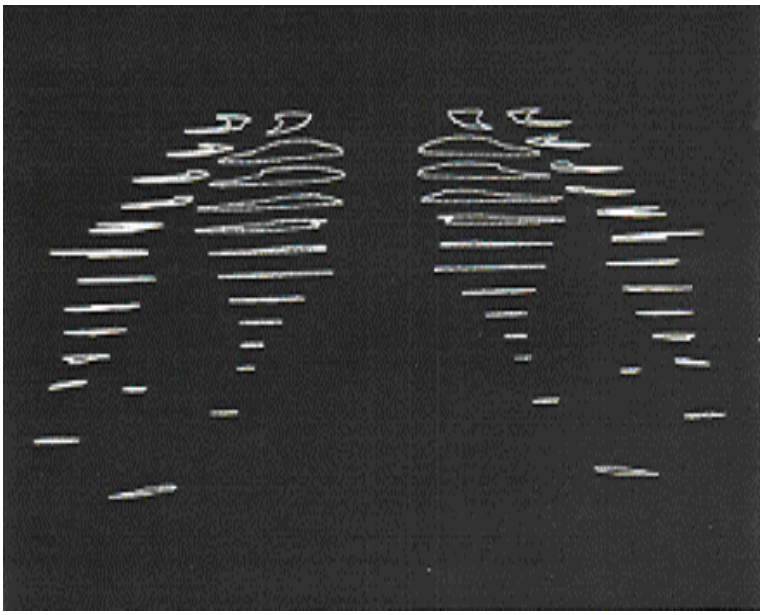


Fig. 8. This figure shows the same structure as in Figure 7. In addition the Pontiform nucleus is present and highlighted and the orientation is approximately the same as in Figure 7

3 The Present – 1990–2002

Many advances took place during this period. Computerized Tomographic Devices have improved both in spatial resolution and signal sensitivity. Just for comparison, in 1989 we used CT scans from GE9800 with spatial resolution of slices 5mm apart and the pixel size was 0.41x 0.41 mm. The PET scanner had ~10mm x 10mm x 10mm resolution. Today the PET scanner has 4mm x 4mm x 4mm. We now have new modalities in magnetic resonance imaging (MRI) devices (1mm x 1mm x 1mm) and their derivatives functional MRI, the fMRI with spatial resolution (2mm x 2mm x 4mm).

All these advances, not only in spatial resolution, but also in signal sensitivities implied the need for better registration algorithms. In this area great progress has been made! It is impossible for me to review all the work and accomplishments, but I must mention a few: Better Atlases: Several groups around the world engaged in the development of Computerized Anatomy (structure) Atlases of the Human Brain for multiple purposes:

1. To develop it as a tool for education
2. Use it for pre-surgical planning
3. Apply the anatomy as a template for segmentation using matching algorithms (examples are: Scaltenbrand-Wharen/Talairach-Tournoux brain atlas [16,17], the Harvard Atlas [10] and others. Better and more sophisticated matching models for structure only (taking into account the tissue/photometric properties).

In our initial model of the atlas, we (Broit and Bajcsy, 1981) considered it as a uniform elastic object that could be "physically" deformed to match the brain image of an individual. Later, Bookstein [5] introduced thin-plate splines model for the same purpose. A different approach was anchored in Theory of Transformations and the use of Grenander's metric pattern theory, as espoused by the group at Brown University and later by M. Miller's group at Johns Hopkins University [12].

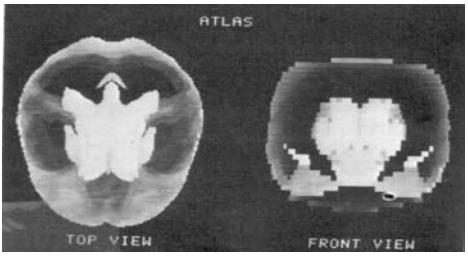


Fig. 9.

The model of Computational Anatomy in their formulation is a Grenader deformable template in which the space of anatomic imagery is in orbit under groups of transformations. There are two types of transformations:

1. Geometry and shape are studied through mappings of the coordinate systems of the anatomies. Variation in the image space is accommodated by introducing group diffeomorphic transformations (invertible differentiable mappings with differentiable inverse) carrying individual elements from one to another.
2. The photometric (gray) values are accommodated by the second transformation. The metric is calculated via equations of motion (change) describing the geodesic connection between the elements.

Yet another approach to account for the variations of anatomic structure and the variability of scanning devices is to use probability theory [8]. The Bayesian methodology facilitates a principled approach to the development of a matching model. This methodology formally embodies the use of prior information. In matching, the prior serves to constrain our mappings by favoring certain configurations, in effect, regularizing the problem. Given the priors, we can then compute the posterior distribution for the unknown mapping conditioned on the current observations. The existence of this distribution makes possible a range of analyses, including the estimation of the variance or reliability of estimated mapping. Needless to say, this approach accounts for both structural variability as well as gray scale/tissue /scanner variability.

4 The Future – 2003 and beyond

Most of the current studies concentrate on processing MRI and/or fMRI images. Positron Emission Tomography (PET) and Single- Photon Emission Computed Tomography (SPECT) are beginning to be used in more clinical settings, combined with

MRI. Comparison between PET data and fMRI data is sought and needs to be established where they are complimentary in detecting functionality.

An exciting and promising imaging modality is in diffusion tensor MRI, originally developed by Peter Basser [14]. The promise of this imaging modality is that it detects directionality and connectivity of neural fiber, hence it has the potential to detect the pathways of neural connections, a long outstanding problem in neural imaging. The challenge for computational anatomy is how to measure similarities for matching such images [2, 20], among other such efforts. See Figure 15 from Similarity Measures for Matching Diffusion Tensor Images, D. Alexander, J. Gee, R. Bajcsy. University of Pennsylvania, Philadelphia, PA, pgs. 93-102.

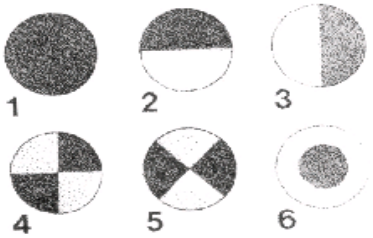


Fig. 10. Pictorial representation of the six "features" or basis functions used as comparators increating the feature space. Dark gray – positive sign; light gray – negative sign

Another challenge is how to recognize and register brain images from patients who have diseases such as Alzheimer's and schizophrenia. These images violate the basic assumption of normal brain images, in which the topology of the brain structure is different. In recognition of these topological differences, several groups developed Disease-Specific Atlases, the Alzheimer's Atlas team [18] and the Schizophrenia Atlas team [13].

Another approach to the diseased atlases has been the use of a deformable probabilistic brain atlas based on random vector field transformations [19]. The algorithm analyzes a reference population of normal scans and automatically generates color-coded probability maps of the anatomy of new subjects. Given a 3-D brain image of a new subject, the algorithm calculates a set of high dimensional volumetric maps (with typically 384 (2) x 256 x 3, approximately 10(8) degrees of freedom) elastically deforming this scan into structural correspondence with other scans, selected one by one from an anatomic image database.

Finally, we would like to utilize the recent advances in gene expression in the brain to identify specific genes that affect some brain activities, either in normal brains or brain diseases. This is difficult to do in vivo thus far, but there is some promise that genes expressed in the brain can sometimes be correlated to genes expressed in peripheral tissues, e.g., blood, which could be used as a diagnostic tool. This in turn can be used for discovering new drug targets.

In order to get a better understanding of these brain activities/processes, one must study animal models which leads us to micro-imaging of various modalities, including spectral imaging in addition to the CT, MRI, PET, SPECT and fMRI imaging. This in turn will be the ultimate challenge for registration processes amongst all of those images.

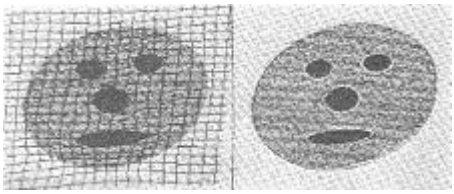


Fig. 11.

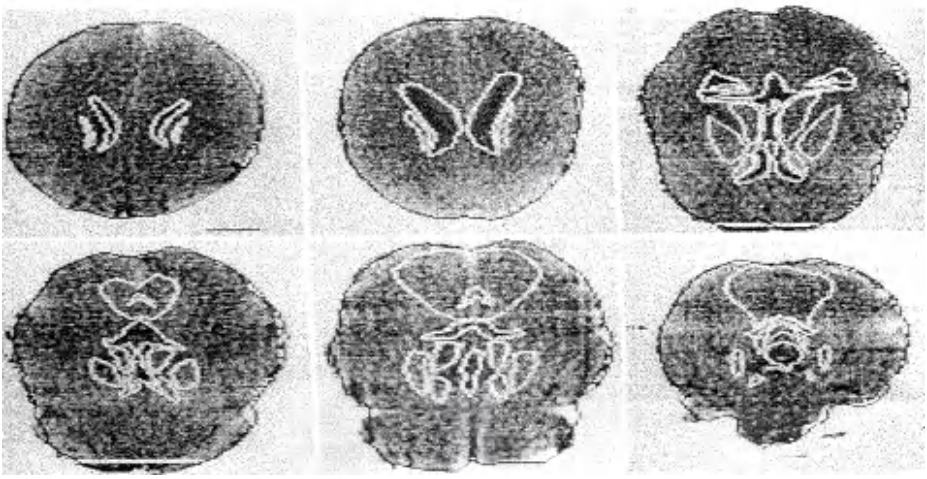


Fig. 12. Results of matching in three dimentions. Shown are the six central slices from a single CT study. A total of 20 stacked atlas slices were used in this matching

5 Conclusions

From this short and certainly incomplete review of the literature of brain imaging and registration, we have shown the progress made over the last twenty plus years - thanks to substantial support from the federal government (NIH) and the ingenuity of the interdisciplinary groups of anatomists, physiologists, computer scientists, physicists, mathematicians and many others.

These advances enable quantitative analyses of these images, accounting for the variability of different scanners, natural variability of human brains, developmental processes of the brain and diseased brains. With the advent of genomic and proteomic research one has the hope of understanding different influences of different genes in the brain and hence capture the onset of brain disease and eventually find the proper drugs to treat that disease.

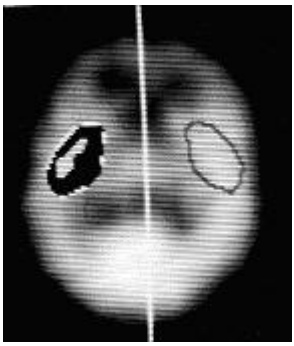


Fig. 13.

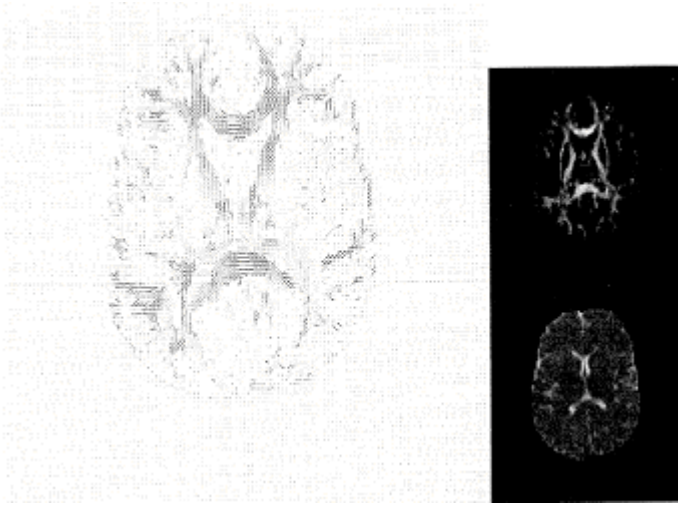


Fig. 14. Images derived from a slice of a DT image. Left: principal DT eigenvector projected into the xy-plane – lighter lines indicate greater z-component. No line is drawn at points where the anisotropy falls below a certain threshold. Top right: lattice anisotropy image. Bottom right: DT trace image

References

1. Adair Taylor, Karp Peter, Stein Alan, Bajcsy Ruzena, Reivich Martin. Computer Assisted Analysis of Tomographic Images of the Brain. Journal of Computer Assisted Tomography, 5(6):929-932;1981.
2. Alexander Daniel, Gee James, Bajcsy Ruzena. Similarity Measures for Matching Diffusion Tensor Images. University of Pennsylvania, Philadelphia, PA, 93-102.
3. Bajcsy Ruzena, Lieberson Robert, Reivich Martin. A Computerized System for the Elastic Matching of Deformed Radiographic Images to Idealized Atlas Images. Journal of Computer Assisted Tomography, 7(4):618-625;1983.

4. Bajcsy Ruzena. Three-Dimensional Analysis and Display of Medical Images. Positron Emission Tomography,119-129;1985.
5. Bookstein FL. Principal warps: Thin-plate splines and the decomposition of deformations. IEEE PAMI,11(6):567-585;1989.
6. Briot C. Optimal Registration of Deformed Images. PhD dissertation, Computer and Information Science Department, University of Pennsylvania, Philadelphia, 1981.
7. Dann Robert, Hoford John, Kovacic Stane, Reivich Martin, Bajcsy Ruzena. Evaluation of Elastic Matching System for Anatomic (CT, MR) and Functional (PET) Cerebral Images. Journal of Computer Assisted Tomography,13(4):603-611;1989.
8. Gee J.C., Le Briquer L., Barillot C., Haynor D.R. Probabilistic Matching of Brain Images. Proceedings of the XIVth International Conference on Information Processing in Medical Imaging, Ile de Berder, France, June 26-30, 1995.
9. Greitz T, Bohm C., Holte S. Eriksson L. A computerized brain atlas: construction, anatomical content and some applications. Journal of Computer Assisted Tomography,15:26-38;1991.
10. Kikinas et al. A Digital Brain Atlas for Surgical Planning, Model Driven Segmentation and Teaching. IEEE Transactions on Visualization and Computer Graphics, http://splweb.bwh.harvard.edu:8000/pages/papers/Anatomy Browser/current/,2(3):September;1996.
11. Mazziotta John, Koslow Stephen. Assessment of Goals and Obstacles in Data Acquisition and Analysis from Emission Tomography: Report of a Series of International Workshops. Journal of Cerebral Blood Flow and Metabolism, 7:S1-S31;1987.
12. Miller Michael, Trouve Alain, Younes Laurent. On the Metrics and Euler-Lagrange Equations of Computational Anatomy. Annual Rev. Biomed. Eng, 4:375-405;2002.
13. Narr KL, Thompson PM, Sharma T, Moussai J, Zoumalan CI, Rayman J, Mazziotta JC, Toga AW (Schizophrenia Atlas Team). Disease-Specific Atlases of the Human Brain. Laboratory of Neuro Imaging, Dept. Neurology, Division of Brain Mapping, UCLA School of Medicine, Los Angeles, CA.
14. Pierpaoli C. and Basser P.J. Toward a Quantitative Assessment of Diffusion Anisotrophy. Magnetic Resonance Medicine, 36:893-906;1996.
15. Schaltenbrand G, Wahren W. Atlas for Stereotaxy of the Human Brain. Chicago, Yearbook, 1977.
16. Talairach J, Szikla G, Tournoux P et al. Atlas d'anatomie stereotaxique du telecephale, Paris, Masson, 1967.
17. Talairach J, Tournoux P. Referentially oriented cerebral MRI Anatomy: an atlas of stereotaxic anatomical correlations for gray and white matter. Stuttgart, Germany:Thieme, 1993.
18. Thompson PM, Mega MS, Dinov ID, Zoumalan CI, Lindshield CJ, Blanton RE, Moussai J, Woods RP, Mazziotta JC, Cummings JL, Toga AW. Disease-Specific Atlases of the Human Brain. Laboratory of Neuro Imaging, Dept. Neurology, Division of Brain Mapping, UCLA School of Medicine, Los Angeles, CA.
19. Thompson PM, Toga AW. Detection, Visualization and Animation of Abnormal Anatomic Structure with a Deformable Probabilistic Brain Atlas Based on Random Vector Field Transformations, Dept. of Neurology, UCLA School of Medicine, Los Angeles, CA.
20. Wiegell Mette, Larsson Henrik, Wedeen Van. Fiber Crossing in Human Brain Depicted with Diffusion Tensor MR Imaging. Radiology, 217(3):897-903;2000.

2D-3D Registration for Interventional Procedures: A Clinical Perspective

Klaus Klingenbeck

Siemens Medical Solutions, Siemensstraße 1, 91301 Forchheim, Germany
klaus.klingenbeck-regn@siemens.com

The clinical application of fluoroscopic imaging systems has undergone a drastic change from diagnostic procedures towards predominant use for real time control of interventions.

Diagnostic imaging modalities, like CT and MR, have been developed from cross sectional imaging to volume imaging with isotropic resolution. Driving elements of this change have been multislice CT-technology and parallel acquisition techniques for MR.

The correspondingly increased speed of data acquisition allowed the 3D-imaging of anatomical volumes with isotropic spatial resolution.

As a necessary consequence also the rendering of the data, viewing and reading has shifted from slices to volumes.

In view of those changes in the field of medical imaging the integration of three-dimensional, morphological as well as functional information into the interventional process has gained significant importance.

This applies to the usage of three-dimensional data for planning the intervention, for guidance as well as for post-interventional control of the therapeutic outcome.

Despite advances of other imaging modalities, fluoroscopy is the method of choice for real time control of interventional procedures and the deployment of interventional devices. Therefore, particularly from this application the need arises for registration of 2D-and 3D-images in combination with the localization of devices in space.

For the general case of 2D-3D registration with data volumes from CT, MR or US corresponding methods did not yet penetrate into clinical use.

However 3D-imaging of high contrast objects with rotating C-arms found a widespread application in neuroradiology.

There it is used to image the vascular tree under selective, arterial injection to identify vascular abnormalities like stenoses, aneurysms or AVMs. Typically the 3D information is used to reliably quantify the extent of such lesions and to identify the feeding vessels for embolization.

Currently the 3D data acquisition with rotating C-arms has several technical limitations, restricting the use to imaging of high contrast structures. Volume imaging with soft tissue resolution – e.g. tumour versus normal tissue – or functional information – e.g. myocardial function and viability – is the domain of CT, MR and nuclear imaging.

The integration of corresponding 3D data into interventional or surgical procedures by 2D-3D registration bears a high potential to improve vascular and nonvascular

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 212–213, 2003.

procedures in terms of reduced procedure times, less radiation exposure to the patient and to the personnel and therapeutic outcomes.

There are numerous examples from tumour embolization to ablative procedures.

An application which entered clinical use recently is magnetic navigation for ablation in cardiac arrhythmia. To develop an “automatic” navigation the use of 3D data sets is a necessary prerequisite, the 2D-3D registration is the link to control under fluoroscopy.

A Comprehensive Approach for Multi-channel Image Registration

G.K. Rohde[1,3], S. Pajevic[2], C. Pierpaoli[1], and P.J. Basser[1]

[1] STBB/LIMB/NICHD
[2] MSCL/CIT, National Institutes of Health, Bethesda, MD, USA
[3] Dept. of Mathematics, University of Maryland, College Park, MD, USA

Abstract. We describe a general framework for multi-channel image registration. A new similarity measure for registering two multi-channel images, each with an arbitrary number of channels, is proposed. Results show that image registration performed based on different channels generates different results. In addition, we show that, when available, the inclusion of multi-channel data in the registration procedure helps produce more accurate results.

1 Introduction

The increasing popularity of new medical imaging modalities capable of producing multivariate data such as Diffusion Tensor MRI (DT-MRI), chemical shift imaging (CSI), phase contrast magnetic resonance angiography, etc., has brought about the need for image registration techniques capable of dealing with multi-channel images. Moreover, existing multi-modal registration techniques (see [1], [2], [3], [4]) can be used to generate multi-channel (multivariate) images from different types of scalar images of a single subject, i.e. conventional MRI and CT scans. All such multi-channel images can benefit from accurate multi-channel image registration procedures. The hope is that the additional information contained in multi-channel images can help render the registration process more accurate. Multiple image channels can be used to characterize each image location more uniquely. Thus unique correspondences between locations of two images are more likely, helping to avoid local optima during registration.

Most existing multi-channel image registration methods are specific to registration of DT-MRI images. Alexander *et al* [5],[6], investigated tensor similarity measures and reorientation approaches to be used when registering DT-MRI data. They reported that a tensor component difference measure combined with the preservation of principal direction tensor reorientation scheme (since tensors contain directional information they must be appropriately reoriented upon spatial transformation of the images) performed best among all options tried [5],[6]. Ruiz-Alzola *et al* [7] demonstrated the feasibility of an approach based on the correlation coefficient of tensor elements as an intensity based similarity function for registration of DT-MRIs. Recently, Guimond *et al* [8] investigated the use of a multivariate version of the daemons algorithm [9]for registering DT-MRI

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 214–223, 2003.

data. Also worth noting is the work by Boes *et al* [10] where the authors used a multivariate mutual information similarity measure to register a scalar image to another set of two images of the same subject.

Here we describe a general framework for registering multi-channel images. A new measure of similarity for two multi-channel images each with an arbitrary number of channels is proposed. We investigate the use of the new method in registering DT-MRI data using both affine and higher order elastic transformations. Experiments using real and simulated datasets show that registration performed based on multiple image channels produces more accurate results than single channel registration.

2 Theory

The goal in image registration is to produce a mapping function $f: \mathbf{x} \rightarrow \mathbf{x}'$ that transforms the spatial coordinates $\mathbf{x}$ of a target image $\mathbf{T}$ to the spatial coordinates $\mathbf{x}'$ of a source image $\mathbf{S}$. For the source and target images to be registered the mapping function f should be chosen in such a way as to maximize some similarity measure (minimize some cost function) between the two images. Mathematically, the image registration problem can be stated as:

$$\max_{f} I(\mathbf{S}(f(\mathbf{x})), \mathbf{T}(\mathbf{x}), f) \tag{1}$$

where $I(\cdot, \cdot)$ represents the chosen similarity measure or cost function. Because of efficiency or accuracy considerations, the multi-channel images $\mathbf{T}$ and $\mathbf{S}$ may be first preprocessed in order to yield best results.

The multi-channel registration problem is described in Figure 1. Multi-channel images are first subjected to a preprocessing step. As described later, this step can either reduce or augment the number of image channels, depending upon the application. Next the images are registered using a multi-channel algorithm soon to be described. Once the spatial transformation that registers the images has been computed, it is applied to the original data and any necessary post processing operation is then performed for final output. The next few sections describe preprocessing approaches, a similarity measure, optimization and approaches for multichannel image registration.

2.1 Preprocessing

The objective of this step is is to transform the original data into images more suitable for registration. When the number of channels is too large, for example, it may be reduced by eliminating any redundant information in the original data using principal component analysis, independent component analysis, or some other kind of dimension reduction or feature extraction technique. Dimension augmentation may be used when the original image channels are not enough to characterize each image region uniquely. Wavelets, for example, could be used to create low-pass and high-pass image channels. Figure 1 depicts this step graphically.

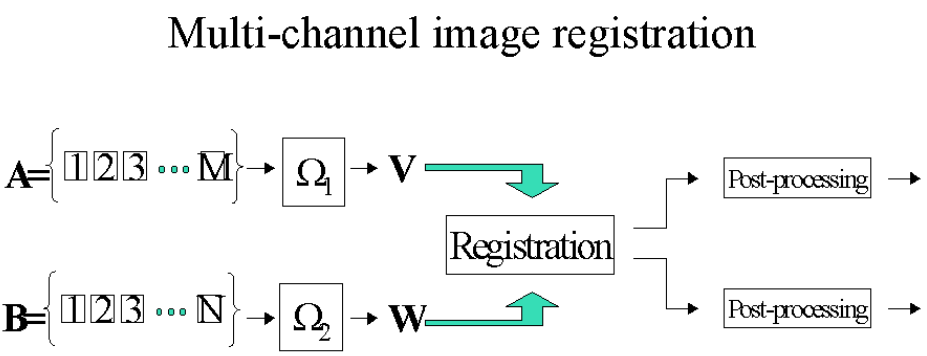


Fig. 1. Description of the possible steps in the multi-channel image registration process. Original images **A** and **B** with M and N channels each are passed through a dimension adjustment operations Ω_1 and Ω_2 producing images **V** and **W**, with P and Q channels, respectively. Images **V** and **W** are then used for registration.

2.2 Similarity Measure

A common methodology used in voxel-based image registration has been to assume that the images are optimally aligned when the statistical dependency between their intensity values is highest. Strategies using least squares, correlation coefficient, and more recently mutual information ([1],[2],[3],[4]), have been used in single or multi-modality registration problems. Here we extend this methodology to the multi-channel registration problem.

Because of the high computational costs associated with estimating the multivariate mutual information between two images with an arbitrary number of channels each, we use the following measure of image similarity:

$$I(\mathbf{S},\mathbf{T}) = \frac{1}{2} log\left(\frac{|\Sigma_{\mathbf{S}}||\Sigma_{\mathbf{T}}|}{|\Sigma|}\right). \tag{2}$$

In equation (2), $|\Sigma|$ is the determinant of joint covariance matrix of multivariate random variables **T** and **S**, given by

$$|\Sigma| = \begin{vmatrix} \Sigma_{\mathbf{T}} & \Sigma_{\mathbf{ST}} \\ \Sigma_{\mathbf{ST}}^T & \Sigma_{\mathbf{S}} \end{vmatrix}. \tag{3}$$

Equation (2) is the multivariate mutual information for normally distributed multivariate random variables **T** and **S** [11]. Note that the computation expressed in equation (2) is much less demanding than estimating the general multivariate mutual information. In particular, when memory requirements are considered, the estimation of the joint probability density function via joint histogram computation for two multi-channel images is prohibitive. In contrast, the memory requirements for computing (2) are negligible, while the most expensive computations involved are the inner products necessary to compute the elements of Σ.

The drawbacks associated with the use of similarity measure (2) stem from the fact that the intensity value of most images are not normally distributed. Therefore it is useful to investigate other interpretations of (2). Using $\Sigma = \mathbf{DPD}$ where $\mathbf{P}$ is a matrix of correlation coefficients and $\mathbf{D}$ is a diagonal matrix of standard deviations, one can show that under fairly weak assumptions (2) may be interpreted as a monotonic function of $|\mathbf{P}|$ (see appendix A). Thus optimizing (2) can be thought of as maximizing the linear correlation coefficient between corresponding channels of each image while minimizing the correlation of channels that do not correspond. We point out that the use of equation (2) to measure image alignment is justified if it can be assumed that there is a linear relationship between the intensity values of the channels of each image being registered.

2.3 Optimization Approaches

The first step in aligning two different images is usually an affine registration step. We use an algorithm similar to what is described in [3] to search for the 12 parameter affine transformation that maximizes Eq. (2). Subsequent elastic registration of the images is performed using the adaptive bases algorithm [12],[13]. The adaptive bases method uses sets of compactly supported radial basis functions to search for the elastic transformation $f(\mathbf{x})$ iteratively through different image resolutions. Note that the optimization of the spatial transformation $f(\mathbf{x})$ here is not crucial and other methods such as described by [14] could be used.

3 Validation Methods

Validation of image registration results, particularly nonrigid ones, is difficult because of the lack of a gold standard. Therefore we resort to indirect measures of alignment. In a first experiment we register a set of 7 images of different healthy volunteers to an 8th image chosen at random. The images used in this experiment are 3D DT-MRI images with three channels each (see description bellow). For comparison purposes, registration of each multi-channel image was performed using one individual image channel at a time as well as all the image channels combined using the multi-channel method previously described; remember, the approach described above allows for registration of images each with an arbitrary number of channels. Thus, by using four different registration methods, we generated 4 sets of 7 multi-channel images each. The first three sets contain multi-channel images that were registered to the template image using a single-channel approach based on channels 1,2, and 3, respectively. The fourth set of images contains the multi-channel images registered to the template using information from all 3 channels simultaneously. In order to determine which set of images was best aligned we use an approach similar to the one described in [15].

We calculate the variance for i^{th} voxel and j^{th} channel, $V(i, j, s)$ across all images in the same set, s. Then we use an average over all voxels to calculate the average variance for each channel $V_c(j, s) = (ni)^{-1} \sum_i V(i, j, s)$, where ni is the number of voxels, and the total average $V_a(s) = (nj)^{-1} \sum_j V_c(j, s)$, where nj

is the number of channels. Note that in order to show relative improvement we report $V_c(j,s)$ as $\frac{V_c(j,s)}{V_c(j,0)}$, where $V_c(j,0)$ is the average variance of the unregistered set. The assumption is that V_c and V_a will be the smallest when the images are optimally aligned.

In a second experiment, a set of 10 3D target multi-channel images was created by applying 10 known transformations to a selected source multi-channel image. Normally distributed noise of variance 8 was added to each generated image whose intensity values ranged from 0 to 255. The transformations were generated using a set of randomly selected Wendland [16] radial basis functions, while the registration algorithm uses one of Wu's [17] radial basis functions. The nonrigid registration algorithm was then used to recover the known transformation using each individual image channel as well as all image channels combined for evaluation of the similarity measure. In addition to multi-channel similarity measure (2) we also used the standard correlation coefficient ρ between all the available image channels to register the images. Thus, in total, we compare 5 methods of registration. Note that due to the limited number of degrees of freedom used in the nonrigid registration algorithm, and the fact that the simulated deformation fields were generated using different basis functions than used in the adaptive bases algorithm, none of the methods can be expected to recover the known transformation without error. Thus we are able to compare the error between the simulated deformations and deformations generated by different registration methods. Note that a similar experimental procedure was used in [18].

4 DT-MRI

All imaging studies were performed with a 1.5 T GE Signa magnet, equipped with a whole-body gradient coil able to produce gradient pulses up to 50mT/m (GE Medical Systems, Milwaukee, WI). The acquisition parameters used are similar to those described in [19].All subjects used in the following experiments were healthy volunteers, both male and female, with ages between 22 and 53. After acquisition the diffusion weighted images were corrected for patient motion and eddy-current induced image distortion using the mutual information based registration approach described in [20].

The multi-channel images were composed of different rotationally invariant scalars extracted from the apparent diffusion tensor model described in [21],[22]. Channel 1 contains the trace of the tensor, channel 2 the relative anisotropy index, and channel 3 the skewness measure. One byte was used to represent the intensity values of each channel. These three specific channels were chosen because the correlation between them is low, implying that each channel contains unique information that is not present in other channels.

5 Results

Figure 2 contains the plot of the multivariate similarity measure (2) as a function of image rotation. The target and source images were generated from a single

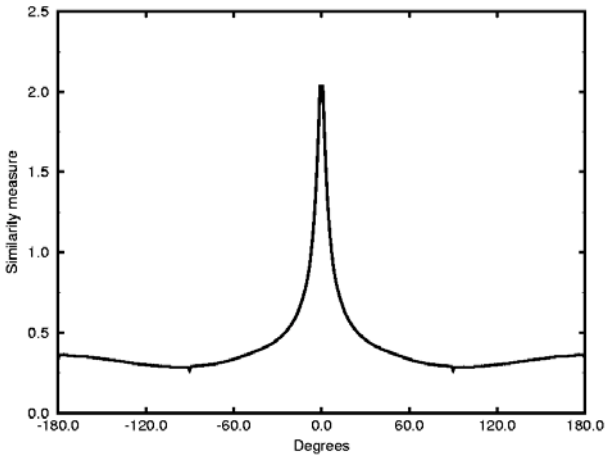


Fig. 2. Plot of multi-channel similarity measure as a function of image rotation.

Table 1. Relative variance for each image channel after affine registration with single and multi-channel approaches.

	$V_c(1,s)$	$V_c(2,s)$	$V_c(3,s)$	$V_a(s)$
ch. 1 registration	0.388	0.430	0.481	0.433
ch. 2 registration	0.507	0.375	0.406	0.430
ch. 3 registration	0.588	0.387	0.404	0.460
multi-channel reg.	0.414	0.397	0.440	0.417

dataset by adding Gaussian noise as described earlier. The images were then rotated with respect to each other and the value of the multi-channel similarity measure (2) is plotted. Note the slight discontinuities around the local extrema are consistent with known effects of image interpolation in registration with mutual information-type cost functions [23].

Table 1 shows the relative improvement in average variance after correction with all approaches mentioned above. Note that only affine transformations were used to align the images in this experiment. When considering the variance of all channels, the multi-channel approach produces best results. The superiority of the multi-channel approach is even more evident when higher order elastic transformations are used (see table2). Figure 3 displays the average of the 7 multi-channel images before and after registration with our approach. The average image of each channel is significantly sharper after registration, indicating good overall alignment of the images.

Table 3 displays the average pixel error for the simulated registration experiments. Evidently, the registration procedure that uses Eq. (2) as similarity measure outperforms all others, including the standard correlation coefficient ρ between all image channels. This suggests that the relationship between unlike, as well as like, channels is important during multi-channel image registration.

Table 2. Relative variance for each image channel after elastic registration with single and multi-channel approaches.

	$V_c(1,s)$	$V_c(2,s)$	$V_c(3,s)$	$V_a(s)$
ch. 1 registration	0.287	0.404	0.474	0.387
ch. 2 registration	0.441	0.354	0.392	0.397
ch. 3 registration	0.630	0.401	0.396	0.473
multi-channel reg.	0.300	0.350	0.395	0.350

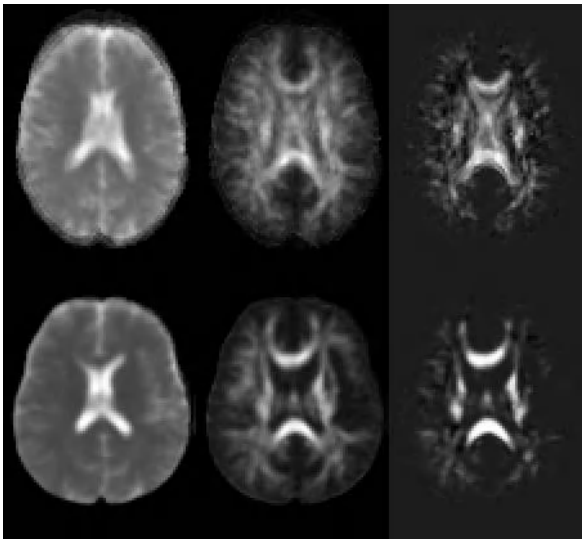


Fig. 3. Average of 7 multi-channel images before and after elastic registration using our multi-channel approach. Each row, from left to right: Trace channel, anisotropy channel, skewness channel.

Table 3. Average pixel error after registration using several methods.

	ch. 1	ch. 2	ch. 3	ρ	Eq. [2]
trial 1	1.55561	1.57457	2.05176	1.57586	1.39538
trial 2	1.86019	1.61742	2.16298	1.84785	1.53049
trial 3	1.64350	1.45177	1.87407	1.68824	1.37878
trial 4	1.60639	1.35870	1.69225	1.49407	1.34622
trial 5	1.63641	1.54283	2.21301	1.74026	1.35014
trial 6	1.54050	1.45536	1.82586	1.54400	1.35805
trial 7	1.68628	1.53352	1.61236	1.59638	1.41817
trial 8	1.66844	1.60115	2.19608	1.74772	1.37936
trial 9	1.76214	1.55805	2.09162	1.74509	1.51251
trial 10	1.76584	1.56465	2.09041	1.74318	1.41244

6 Conclusions

Here we have described a general framework for registering a multi-channel image with an arbitrary number of channels to another image also with an arbitrary number of channels using a multivariate correlation approach. Our approach not only considers the similarity between like-channels of each image, but also the similarity between channels that do not correspond. Although we have used diffusion tensor images of the human brain for validation, we believe our method to be applicable to many other multi-channel registration problems.

Results show that image alignment based on different image channels can lead to different registrations. This is true especially in the nonrigid case, but also for the affine one. When considering image alignment across all channels, results indicate that, even for the affine case, the use of all channels simultaneously using the similarity measure defined in Eq. [2] during registration is superior to any of the other methods investigated.

Appendix

We show that under fairly weak assumptions Eq. (2) is a monotonic function with respect to $|\mathbf{P}|$. Using $\Sigma = \mathbf{DPD}$, where $\mathbf{P}$ is a matrix of correlation coefficients and $\mathbf{D}$ is a diagonal matrix of standard deviations,

$$\mathbf{D} = \begin{pmatrix} \mathbf{D_T} & \mathbf{0} \\ \mathbf{0} & \mathbf{D_S} \end{pmatrix}, \tag{4}$$

Eq. (2) becomes

$$I(\mathbf{S}, \mathbf{T}) = \frac{1}{2} log \left(\frac{|\Sigma_\mathbf{S}||\Sigma_\mathbf{T}|}{|\mathbf{D}_\mathbf{S}^2||\mathbf{D}_\mathbf{T}^2||\mathbf{P}|} \right). \tag{5}$$

Note that $\frac{|\Sigma_\mathbf{T}|}{|\mathbf{D}_\mathbf{T}^2|}$ is constant in (5) because the target image is static throughout the registration procedure. If we assume that spatial transformations f do not change the covariance between the image channels of $\mathbf{S}$ significantly $\frac{|\Sigma_\mathbf{s}|}{|\mathbf{D}_\mathbf{s}^2|}$ is also constant and (5) is a monotonic function with respect to $|\mathbf{P}|$. This assumption is reasonable for f modifies all image channels identically. Therefore one can expect the relationship between the image channels not to change significantly. Note also that if the image channels of $\mathbf{S}$ (and of $\mathbf{T}$) are uncorrelated further simplification is possible and (5) becomes $I(\mathbf{S}, \mathbf{T}) = -0.5 * log(|\mathbf{P}|)$.

Lastly, note that the multi-channel similarity measure (2) reduces to a monotonic function of the scalar correlation coefficient when the images have one channel each. In this case

$$|\Sigma| = \begin{pmatrix} Cov(\mathbf{S}, \mathbf{S}) & Cov(\mathbf{S}, \mathbf{T}) \\ Cov(\mathbf{T}, \mathbf{S}) & Cov(\mathbf{T}, \mathbf{T}) \end{pmatrix} \tag{6}$$

and (2) reduces to $I(\mathbf{S}, \mathbf{T}) = -0.5 * log(1-\rho)$, where ρ is the correlation coefficient between the intensity values of the images.

References

1. Viola, P., Wells, III, W. M.: Alignment by maximization of mutual information. Proc. Int. Conf. Computer Vision, Boston (1995) 16–23
2. Wells, III, W. M., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R.: Multi-modal volume registration by maximization of mutual information. Medical Image Analysis. **1** (1996) 35–51
3. Maes, F., Collignon, A., Vandermeulen, D., Marcha, G., Suetens, P.: Multimodality image registration by maximization of mutual information. IEEE Transactions on Medical Imaging **16** (1997) 187–198
4. Studholme, C., Hill, D. L. J., Hawkes, D. J.: Automated trhee-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures. Medical Physics **24** (1997) 25–35
5. Alexander, D. C., Gee, J. C., Bajcsy, R: Transformations of and similarity measures for diffusion tensor MRIs. Workshop on Biomedical Image Registration, at CAIP (1999)
6. Alexander, D. C., Pierpaoli, C., Basser, P. J, Gee, J. C: Spatial Transformations of Diffusion Tensor Magnetic Resonance Images. IEEE Transactions on Medical Imaging **20** (2001) 1131–1139
7. Ruiz-Alzola, J., Westin, C.-F, Warfield, S. K., Alberola, C., Maier, S., Kinikis, R: Nonrigid registration of 3D tensor medical data. Medical Image Analysis **6** (2002) 143–161
8. Guimond, A., Guttmann, C. R. G., Warfield, S. K., Westin, C.-F: Deformable registration of DT-MRI data based on transformation invariant tensor characteristics. Proc. ISBI'02 (2002)
9. Thirion, J.-P: Image matching as a diffusion process: an analogy with Maxwell's daemons. Medical Image Analysis, **2** (1998) 243–260
10. Boes, J. L., Meyer, C. R.: Multi-variate mutual information for registration. Proc. MICCAI'99 (1999) 606–612
11. Salomon Kullback: Information Theory and Statistics. Dover Publications, Mineola, New York (1968)
12. Rohde, G. K., Aldroubi, A., Dawant, B. M.: The adaptive bases algorithm for nonrigid image registration. Proc. SPIE Medical Imaging, San Diego, USA (2002)
13. Rohde, G. K.: The adaptive grid registration algorithm: a new spline modeling approach for automatic intensity based nonrigid registration. Masters thesis, Dept. Electrical Engineering, Vanderbilt University (2001)
14. Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L. G., Leach, M. O., Hawkes, D. J.: Nonrigid registration using free-form deformations: application to breast MR images. IEEE Transactions on Medical Imaging **18** (1999) 712–721
15. Studholme, C., Cardenas, V. A., Weiner, M. W.: Multiscale image and multiscale deformation of brain anatomy for building average brain atlases. Proc. SPIE Medical Imaging, Sand Diego, USA (2001)
16. Wendland, H.: Piecewise polynomial, positive definite and compactly supported radial basis functions of minimal degree. Advances in Computational Mathematics, **4** (1995) 389–396
17. Schaback, R., Creating surfaces from scattered data using radial basis functions. in Mathematical Models for Curves and Surfaces. Vanderbilt University Press, Nashville, TN (1995)

18. Rueckert, D., Clarkson, M. J., Hill, D., L., G., Hawkes, D., J.: Non-rigid registration using higher-order mutual information. Proce SPIE Medical Imaging, San Diego, 2000.
19. Pierpaoli, C., Jezzard, P., Basser, P. J, Barbett, A., DiChiro, G.: Diffusion tensor imaging of the human brain. Radiology **201** (1996) 637–648
20. Rohde, G. K., Barnett, A., S., Basser, P., J., Marenco, S., Pierpaoli, C.: A comprehensive approach for correcting motion and distortion in diffusion weighted MRI. Proc 21st Southern Biomedical Engineering Conference, Bethesda, MD, (2002)
21. Basser, P. J., Mattiello, J., Lebihan, D.: MR Diffusion Tensor and Imaging. Biophysical Journal **66** (1994) 259–267
22. Basser, P. J.: New histological and physiological stains derived from diffusion-tensor MR images. Annals of the New York Academy of Sciences **820** (1997) 123–138
23. Pluim, J., Maintz, J.,Viergever, M.: Interpolation artifacts in mutual information-based image registration. Computer Vision and Image Understanding **77** (2000) 211–232

Free Form Deformations Guided by Gradient Vector Flow: A Surface Registration Method in Thoracic and Abdominal PET-CT Applications

Oscar Camara, Gaspar Delso, and Isabelle Bloch

Ecole Nationale Supérieure des Télécommunications
Département TSI, CNRS URA 820, 46 rue Barrault
75634 Paris Cedex 13, France
{Oscar.Camara,Gaspar.Delso,Isabelle.Bloch}@enst.fr

Abstract. A nonlinear surface registration algorithm of thoracic/abdominal structures segmented from CT and PET volumes is presented. The aim of this work is to develop a method that can provide an initial estimate of the elastic deformation between the images, so that MI-based techniques can be successfully applied. To perform the matching, a B-spline Free Form Deformation (FFD) model has been chosen. Hierarchical structure segmentation and rigid registration are applied to initialize the nonlinear surface registration phase. Two different approaches to optimize the warp are tested: an iterative gradient descent technique based on local gradient estimations over the grid of control points; and an original optimization based on Gradient Vector Flow (GVF) computed on the CT image. Finally, we evaluate our results, using an Iterative Closest Point (ICP) rigid registration algorithm as a reference to compare both approaches.

1 Introduction

Image registration is an image processing tool that can significantly improve clinical decisions in oncology applications. In this field, the need to combine anatomical and functional images is increasing, due to the development of new acquisition devices and methods that provide complementary information. Positron Emission Tomography (PET) acquisitions provide rich functional information for diagnostic and therapeutic follow-up of cancer, but make it difficult to precisely locate the tumors with respect to the surrounding anatomy. On the other hand, Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) give access to this anatomical information, but do not offer sufficient data to state the lesion malignancy. Therefore, combining information from these two modalities would have a significant impact on improving medical decisions for diagnosis, therapy and treatment planning [1]. To perform such a combination, a registration step is needed in order to grant a good correspondence between the anatomical structures contained in the images.

In thoracic/abdominal oncology, the main problem to be solved to achieve the registration is to deal with the severe deformations undergone by the anatomical

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 224–233, 2003.

structures of interest due to the different acquisition protocols involved, the elastic nature of the imaged regions and the patient's own metabolic activity. For this reason, linear transformations (rigid or affine) cannot be employed to model the deformation between both image modalities.

Several nonlinear transformations can be found in the image processing literature, such as thin-plate splines [2], elastic [3] and viscous fluid [4] models. However, our application imposes certain conditions: a high enough number of degrees of freedom to cope with large deformations; a limited computational cost to be able to use it in clinical routine; and the impossibility of employing external landmarks to help the registration procedure. B-Spline Free Form Deformations (FFD), introduced by Sederberg et al. [5], is a parametric technique which provides flexible nonlinear transformations which fulfill our needs.

FFD combined with Mutual Information (MI) criterion have been successfully used in different medical imaging applications, such as pre-and post contrast MR mammogram registration [6], brain registration [7] or cardiac segmentation [8]. Most of these applications work with monomodality data, where relationship between corresponding intensities in both images to register is simpler than in multimodality applications. That is not our case, as thoracic/abdominal PET and CT images show larger intensity differences between corresponding structures and a big deal of noise and artifacts, that can cause a MI-based algorithm to converge towards non optimal registration results.

One way to avoid this problem is to initialize these deformations close enough to the final solution, thus reducing the risk of falling into local minima. We have proved in [9] how this initialization can improve MI-based nonlinear registration results, this posterior phase being considered as a refinement step of the initialization results, capable of correcting any errors the segmentation might have induced and improving the registration of those regions distant from the segmented structures. In this article, we present an original nonlinear registration method between homologous structures segmented in both PET and CT images, to initialize the grey-level FFD-based multimodality registration. The main contribution of this paper, concerning the registration domain, is the use of a Gradient Vector Flow vector field computed on the CT structures to guide the FFD model applied on the PET objects.

The article is organized as follows. First, we describe the preprocessing stage including structure segmentation and the initial rigid registration. In Section 3, we present both FFD-based approaches to perform nonlinear surface registration. Comparative evaluation of the methods is discussed in Section 4 and, finally, conclusions are presented.

2 Preprocessing

2.1 Structure Segmentation

The registration step requires the segmentation of some common structures in both images. It is important to note that the quality of the segmentation does not need to be perfect, as the target is just to provide an initial estimate of the

Fig. 1. Left: 3D rendering of segmented structures in CT image superimposed on a coronal slice. Right: 3D rendering of segmented structures in PET image superimposed on a coronal slice.

transformation between these structures and errors induced by this stage will be corrected by the posterior grey-level registration phase. Working with several clinicians, we have determined a list of structures that can always be found in thoracic/abdominal PET and CT images. For the time being, we have worked with the three most relevant organs: lungs, liver and kidneys. Segmentation of these anatomical features is not a trivial task, especially in PET scans, where gradients are weak and sometimes the poor image quality does not allow to recognize even visually the different structures.

Due to the above reasons, we propose a hierarchical procedure: the extraction of a given structure will be driven by information derived from a simpler one. This information is composed both of spatial constraints and relationships inferred from the previously segmented structures, expressed by means of Regions Of Interest (ROI) in which the search for new structures will take place. The order in which structures are segmented is: body contours, lungs, kidneys and liver. Some results of the final segmentation are shown in Figure 1.

Each structure is processed in two different stages: a first rough segmentation and a refinement of the result. Rough segmentation is basically composed of automatic thresholding and mathematical morphology operations in the ROI defined by previously segmented objects. Some properties of the images and of the anatomy are exploited in order to make this phase more robust, especially in PET images:

– PET lungs: PET acquisition often provides emission and transmission images, the lungs being easier to segment in the latter. We perform lung segmentation in the transmission image to initialize the segmentation in the emission image, thus improving the quality and the robustness of the results.

- Kidneys: These structures are segmented by selecting the two most symmetrical components with respect to the body's symmetry plane. The symmetry plane computation and the selection of symmetrical regions are made using an algorithm and a symmetry measure proposed in [10].
- Liver: Often, the liver is much easier to segment in one modality than in the other. In those cases, we roughly register segmented structures from one modality to the other and dilate them to produce a ROI, which restricts the search in the more difficult images.

The results obtained from the previous stage cannot be considered as a final result. In particular, the lack of a regularization term may lead to imperfect boundaries. These problems are overcome using a 3D simplex mesh deformable model [11].

2.2 Rigid Surface Registration

Once the organs are conveniently segmented, they can be easily used to automatically establish a first approximation of the registration. It includes a rigid motion, independent scaling along the three axes and cropping out those parts of the volumes without a correspondence or that have no interest for our application.

3 Nonlinear Surface Registration

The registration of 3D surfaces has been extensively treated in various fields of image processing and computer vision. Interesting reviews of these techniques can be found in [12] [13].

One of the most successful methods is the Iterative Closest Point (ICP), proposed by Besl et al. [14], and it has been widely used with good results in several applications. This algorithm iteratively tunes a rigid transformation to minimize the distance between two point (line, triangle, ...) sets whose correspondences have been determined by proximity. In previous works [9], we have employed the distance map computed for this method to perform an estimation of the nonlinear transformation between the surfaces of each pair of organs. This was used to initialize the positions of the control points of the FFD of the grey-level MI registration stage. Nevertheless, this method presented some problems, such as erroneous deformations around those structures without an obvious correspondence. Furthermore, the interaction with FFD-based grey-level registration was solved in a way where some information could be lost.

For these reasons, we have chosen to apply to this surface registration stage the same FFD deformation framework used in the grey-level registration stage, simplifying the interaction between both phases. In this section, after a brief introduction to the FFD model, we propose two FFD-based approaches to achieve nonlinear surface registration. The first method employs a Root Mean Square (RMS) criterion as a similarity measure, computing a local estimation of the gradient at each control point of the grid, and using an iterative gradient descent

procedure. The second method computes a Gradient Vector Flow (GVF) field over CT segmented structures to guide the optimization of the control points.

3.1 Free Form Deformation Model

In this technique, deformations of the object are achieved by tuning an underlying mesh of control points, which in the present application has been set to $10 \times 10 \times 10$ effective nodes. The control point displacements are then interpolated to obtain a smooth and continuous C^2 transformation. A B-Spline based FFD can be written as a 3D tensor product of one-dimensional cubic B-Splines. Let Φ denote an uniformly spaced grid of $n_x \times n_y \times n_z$ control points $\phi_{i,j,k}$ with a spacing of δ, where $-1 \leq i < n_x - 1, -1 \leq j < n_y - 1, -1 \leq k < n_z - 1$. Then, the nonlinear displacement field for each image point *x,y,z* is computed:

$$d(x, y, z) = \sum_{l=0}^{3} \sum_{m=0}^{3} \sum_{n=0}^{3} \theta_l(u)\theta_m(v)\theta_n(w)\phi_{i+l,j+m,k+n} \tag{1}$$

Here, *i*, *j*, and *k* denote the indices of the control point cell containing *(x,y,z)*, and *u*, *v*, and *w* are the relative positions of *(x,y,z)* in the three dimensions, θ_0 through θ_3 being 1D cubic B-Splines.

3.2 Free Form Deformations Guided by Root Mean Square (RMS-FFD)

We have to chose a similarity criterion to control the optimization of the FFD grid of control points. Segmented images have a linear intensity relation, thus the simpler way is to employ the Root Mean Square (RMS) difference of corresponding pixels intensities. We could use other simpler similarity measures but we have chosen the RMS in order to deal with several structures in each image at the same time, as it is planned in future works.

The optimization procedure is based on an iterative gradient descent technique over the entire grid of control points. At each iteration, we compute a finite differences local gradient estimation for each control point. Furthermore, a local spring force regularization term, pulling each node towards the centroid of its neighbouring nodes, has been added to avoid overfitting and to prevent the nodes from intersecting, which could lead to unwanted alterations of the structure topology. Results of this method applied to the lungs and liver are shown in Figure 2, where the good performance of this nonlinear registration algorithm can be appreciated.

3.3 Free Form Deformations Guided by Gradient Vector Flow (GVF-FFD)

A drawback of the RMS-FFD method is the need to optimize all the control points with the local gradient estimation computed at each iteration. Multiresolution approaches accelerate the convergence of the algorithm, but the gradient estimation remains a problem in terms of computation time.

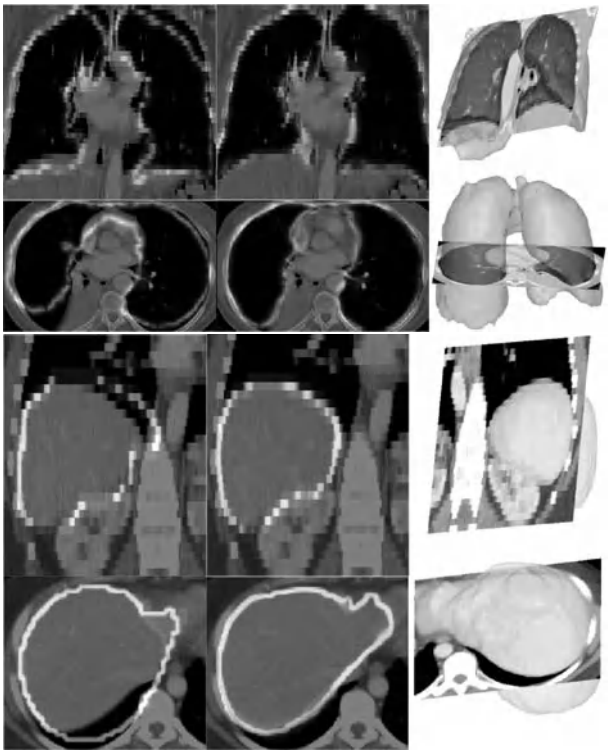


Fig. 2. Original coronal and axial CT slices are superimposed with: Left: contour of rigidly registered PET structures; Center: contour of PET structures nonlinearly registered with RMS-FFD method; Right: 3D rendering PET structures nonlinearly registered with RMS-FFD method. 3D images shown in this article have been visualized using the Anatomist software (www.anatomist.info), developed at S.H.F.J, Orsay.

We propose an original way to speed-up the optimization of the deformation, computing a vector field over the target structures (in our case, those segmented from the CT images) to guide the tuning of the Free Form Deformation applied to the source structures (in our case, those segmented from the PET images).

Gradient Vector Flow, proposed by Xu et al. [15], is an elegant method to obtain this vector field, giving us at each image point the direction towards the target anatomical feature. This technique is usually used to guide deformable models in segmentation applications, but to our knowledge, it has not been used before to control FFD in registration algorithms.

A GVF field $\mathbf{v}$ is defined as the equilibrium solution of the following vector diffusion equation:

$$\mathbf{u}_t = g(|\nabla f|)\nabla^2\mathbf{u} - h(|\nabla f|)(\mathbf{u} - \nabla f) \tag{2}$$

$$\mathbf{u}(x, 0) = \nabla f(x) \tag{3}$$

where $f(x)$ has been obtained applying a Canny edge detector to the target image. The first term at the right side of equation (2) is called *smoothing term*, and tends to uniformize the resulting vector field. The second term is the *data term*, and drives the vector field u towards the $\bigtriangledown f$ computed from the data. $g()$ and $h()$ are weighting functions that apply respectively to the smoothing and data term. A common choice for these is:

$$g(|\bigtriangledown f|) = e^{-\left(\frac{|\bigtriangledown f|}{K}\right)^2}, \qquad h(|\bigtriangledown f|) = 1 - g(|\bigtriangledown f|) \tag{4}$$

This formulation allows for the gradient vector diffusion only where there are no strong edges, thus preventing the smoothing effect from averaging close opposing gradients, which could lead to a gap in the contour through which our model could leak.

After the computation of the GVF vector field, we have for each point belonging to a PET structure the right direction to evolve towards its corresponding CT structure. Thus, for each grid control point, we find the contour points of the PET structures under its influence and we observe the value of GVF vector field for these points. These values are weighted according to their grid control point and the mean of the resulting vectors is taken as the optimal control point displacement direction. At the end of each iteration, a local spring regularization term is applied like in the RMS-FFD algorithm.

The convergence of the algorithm depends on the quality of the computed vector field. Moreover, control points tend to oscillate in the vicinity of CT structure boundaries. Therefore, several constraints over the control points have been implemented to stop them when they start to oscillate. We have introduced the procedure in a multi-step framework to help the algorithm to cope with large deformations in the first iterations, leaving more local ones at the end. Results of this method applied to the lungs and the liver structures are shown in Figure 3. We can appreciate that we obtain lower quality deformations with respect to the RMS-FFD method. On the other hand, this algorithm converges much faster towards correct transformations, as we show in Section 4.

4 Comparison of the Methods

In this section, we compare quantitatively the quality and computational cost the two proposed FFD-methods compared to the ICP algorithm. We have applied these techniques over a database composed of 15 pairs of deformable structures, such as lungs, liver or kidneys, previously segmented from CT and PET images.

To evaluate quality other than by visual inspection of 3D volumes, we have computed two quantitative measures: an overlap measure directly applied on the segmented structures and the Mutual Information (MI) applied on the grey-level intensity images. The reason of using MI instead of RMS directly over segmented structures is to take into account the possible errors induced by the segmentation phase.

The overlap measure consists of the quotient between intersection and union among segmented structures, and is equal to 1 if total superimposition is

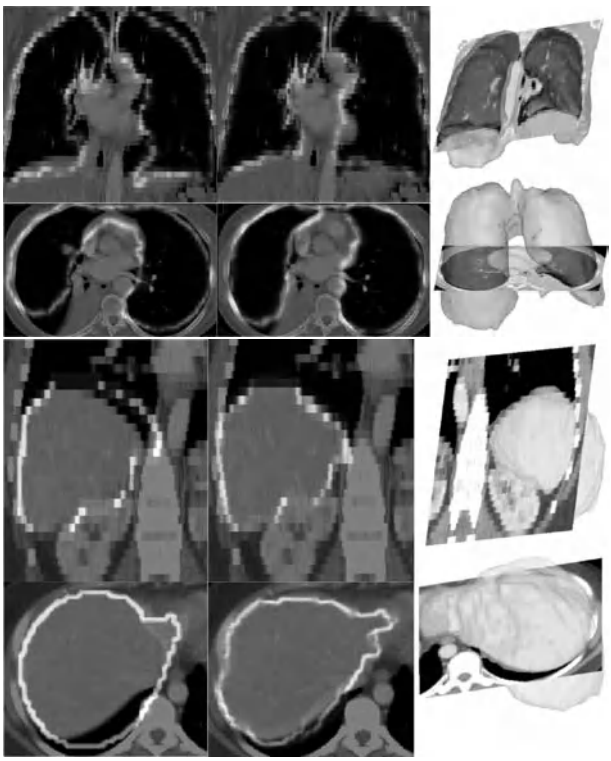


Fig. 3. Original coronal and axial CT slices are superimposed with: Left: contour of rigidly registered PET structures; Center: contour of PET structures nonlinearly registered with GVF-FFD method; Right: 3D rendering PET structures nonlinearly registered with GVF-FFD method.

achieved. Mutual Information (MI) criterion, introduced by Viola [16], is a powerful tool for multimodal image registration with nonlinear intensity relation. It expresses how much information from an image I is contained in another image J. Therefore, MI will be maximal if the images are geometrically aligned. Thus, we compute MI between grey level CT image and grey-level PET image, registered by applying the FFD computed over the surfaces. Table values of the overlap measure and MI are completed with the improvement percentage respect to the ICP results.

Furthermore, we have analyzed the computational costs of each method. We have normalized the results by the dimensions of each image, to be able to compare directly method performances, independently of the registered structures. Thus, table values of time are in μs./pixel. Quantitative results and computational costs are summarized in Table 1.

The overlap measure and MI confirm the visual results, that is, RMS-FFD is the method which obtains the best deformations, while GVF-FFD still clearly surpasses rigid registration methods. It is worth noticing how the overlap of the

Table 1. Evaluation results.

Method	**ICP**	**RMS-FFD**	**GVF-FFD**
Overlap(value/%)	0.6586/100	0.9030/137.095	0.8275/122.081
MI(value/%)	0.1888/100	0.2592/137.286	0.2486/131.697
Time	6.60723	699.365	52.610

segmented structures never reaches 100 %; this is due to the inherent regularization of the chosen elastic deformation model, which prevents the finer differences from being removed. As a matter of fact this is quite convenient, because those differences are usually due to the imaging modalities involved rather than to misregistration, and must be preserved as they are.

We have observed that RMS-FFD works better than GVF-FFD in regions with very small deformations. This is due to the trade-off in the computation of the GVF vector field between the rejection of undesired structures, such as bronquia in the case of lungs, and the capacity to cope with local deformations.

On the other hand, GVF-FFD shows better performance in terms of rapidity of convergence, while RMS-FFD remains a high-computational cost algorithm, due to the computation of local gradient estimations in each iteration. We have tried initializing RMS-FFD with the solution furnished by the GVF-FFD, obtaining even better quality results than using only RMS-FFD, while having acceptable convergence times.

5 Conclusions

We have proposed an original nonlinear surface registration framework based on a FFD model applied to segmented structures from thoracic/abdominal CT and PET images, aimed at initializing a posterior grey-level MI nonlinear registration algorithm. Two different approaches to optimize the deformation, RMS-FFD and GVF-FFD have been presented. A quantitative evaluation of these methods compared to the classical ICP algorithm has been performed. RMS-FFD has proved to be the method providing best results in terms of registration quality, at the expense of a high computational cost. On the other hand, GVF-FFD, furnishing slightly less accurate deformations, converges towards the solution in a much shorter time.

Acknowledgment

The authors would like to thank Dr. Hervé Foehrenbach, Dr. Pierre Rigo, Dr. Marchandise for their contribution to this project. This work was partially supported by the French Ministry for Research (grant number 01B0267).

References

1. H. N. Wagner, "Fused Image Tomography: An Integrating Force," *Nuclear Medicine*, vol. 40, no. 8, pp. 13N–32N, 1999.
2. F. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 6, pp. 567–585, 1989.
3. R. Bajcsy and S. Kovacic, "Multiresolution elastic matching," *Computer Vision, Graphics and Image Processing*, vol. 46, pp. 1–21, 1989.
4. M. Bro-Nielsen and C. Gramkow, "Fast fluid registration of medical images," in *VBC*, Hamburg, Germany, Sept. 1996, pp. 267–276.
5. T. Sederberg and S. Parry, "Free form deformation of solid geometric models," in *SIGGRAPH'86*, vol. 20, Dallas, USA, August 1986, pp. 151–160.
6. D. Rueckert, I. Somoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes, "Nonrigid Registration Using Free-Form Deformations: Applications to Breast MR Images," *IEEE Transactions on Medical Imaging*, vol. 18, no. 8, pp. 712–721, 1999.
7. T. Hartkens, D. Hill, A. Castellano-Smith, D. Hawkes, C. M. Jr., A. Martin, W. Hall, H. Liu, and C. Truwit, "Using points and surfaces to improve voxel-based non-rigid registration," in *MICCAI*, 2002, pp. 565–572.
8. J. Lotjonen, "Segmentation of mr images using deformable models: Application to cardiac images," *International Journal of Bioelectromagnetism*, vol. 3, no. 2, pp. 37–45, 2001.
9. O. Camara, G. Delso, and I. Bloch, "Evaluation of a thoracic elastic registration method using anatomical constraints in oncology," in *EMBS-BMES'02*, Houston, USA, October 2002.
10. O. Colliot, I. Bloch, and A. Tuzikov, "Characterization of approximate plane symmetries for 3D fuzzy objects," in *IPMU*, vol. 3, Annecy, France, July 2002, pp. 1749–1756.
11. H. Delingette, "General object reconstruction based on simplex meshes," *International Journal of Computer Vision*, vol. 32, no. 2, pp. 111–146, 1999.
12. M. Audette, F. Ferrie, and T. Peters, "An algorithmic overview of surface registration techniques for medical imaging," *Medical Image Analysis*, vol. 4, pp. 201–217, 2000.
13. J. B. A. Maintz and M. A. Viergever, "A Survey of Medical Image Registration," *Medical Image Analysis*, vol. 2, no. 1, pp. 1–36, 1998.
14. P. Besl and N. McKay, "A method for registration of 3D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 14, pp. 239–256, 1992.
15. C. Xu, "Deformable models with application to human cerebral cortex reconstruction in magnetic resonance images," Ph.D. dissertation, Johns Hopkins University, 2000.
16. P. Viola, "Alignment by maximization of mutual information," Ph.D. dissertation, MIT, Cambridge, Ma, 1995.

High-Dimensional Multi-modal Image Registration

Peter J. Lorenzen[1] and Sarang C. Joshi[1,2]

[1] Department of Computer Science
University of North Carolina at Chapel Hill
[2] Department of Radiation Oncology
University of North Carolina at Chapel Hill

Abstract. This paper presents a Bayesian framework for generating inter-subject high-dimensional transformations between two multi-modal image sets of the brain. In this framework, the estimated transformations are generated by using the maximal information about the underlying neuroanatomy present in each of the different modalities. This modality independent registration framework is achieved by using the Bayesian paradigm and jointly estimating the posterior densities associated with the multi-modal image sets and the high dimensional registration transformation mapping the two subjects. The methods presented do not assume that the same modalities were used to image the two subjects. To maximally use the information present in all the modalities, relative entropy (or Kullback Leibler divergence) between the estimated posteriors is minimized to estimate the registration. The high-dimensional registration is constrained to be diffeomorphic by using the large deformation fluid formulation.
We also show that the general framework optimally unifies the processes of segmentation and registration of human brain anatomy. Given a set of multi-modal intensity images of an individual subject and a geometric atlas prior, we jointly estimate the posterior that represents the structure of that subject's neuroanatomy and the transformation that maps the space of the atlas prior to the space of the estimated posterior, capturing high-dimensional local variability.

Keywords: Multi-modal image registration, image segmentation, information theory, medical image analysis, computational anatomy.

1 Introduction

Modern imaging techniques provide an array of imaging modalities which enable the acquisition of complementary information representing underlying neuroanatomy. Certain imaging sensors, such as MRI, produce inherently registered multi-modal images. In the case were the images are derived from disparate sensors with differing geometry (e.g. CT and MRI), rigid registration is required to bring the modalities into correspondence. Multi-modal rigid image registration has received much attention and a thorough review and comparison of methods is presented in [20].

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 234–243, 2003.

Inter-subject high-dimensional image registration has proven to be a powerful tool in understanding brain anatomy [5,2,7]. The book *Brain Warping* [17] discusses the application of high-dimensional image registration to the understanding of neuroanatomy.

Although inter-subject high-dimensional image registration has received much attention [15,10,6,16], to our knowledge, little attention has been given to using multi-modal image sets of subjects to estimate the registration transformation. Most image registration techniques use a single imaging modality for the generation of high-dimensional transformations between images of two subjects. An algorithm for cross modality inter-subject high-dimensional image registration that map image intensities from one modality to another has been proposed [8]. This algorithm only works with a single image from each subject.

In this paper, we present a framework based on the Bayesian paradigm that enables us to generate inter-subject high-dimensional transformations given two acquired sets of multi-modal images. We assume that each subject has associated with it a set of imaging modalities (e.g. PET, MRI, CT, etc.). We develop an algorithm in which the maximal amount of information present in the different modalities is used to generate the transformation. Throughout this paper we assume that, for each subject, the multi-modal images of that subject have been registered to one another.

The framework is based on the assumption that human brain anatomy consists of finitely enumerable structures such as grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF). These structures present with varying radiometric intensity values across disparate image modalities. Given multi-modal image sets representing two studies, we jointly estimate, for each subject, the posterior distributions associated with each of the structures along with the diffeomorphic high-dimensional registration map that relates the coordinate spaces of the two subjects. The Kullback-Leibler divergence is used as a metric for the posterior densities to estimate the transformation. The use of the posterior probability densities provides an image intensity independent approach to image registration.

In this paper, we also show that the framework described above unifies the processes of segmentation and registration of human brain anatomy. Given a set of multi-modal intensity images representing an individual subject, and a geometric atlas prior, the same framework can be used to jointly estimate the posterior that represents the structure of that subject's neuroanatomy and the transformation that maps the space of the atlas prior to the space of the estimated posterior. We are not the first to consider using an atlas prior in automatic segmentation of multi-channel images using non-rigid registration. A method for incorporating non-rigid registration into an expectation maximization algorithm to segment magnetic resonance images is presented in [14]. With respect to the non-rigid registration, our method fundamentally differs from the preceding approach in that the registration and segmentation are concurrently estimated.

The remainder of the paper is organized in the following manner. Section 2 presents the high-dimensional multi-modal image registration framework de-

scribing the posterior and transformation estimation. In Section 3, this framework is applied to the problem of atlas-based image registration and segmentation. In Section 4, we evaluate the performance of the segmentation algorithm.

2 High-Dimension Multi-modal Image Registration

We first consider the problem of registering two acquired sets of multi-modal images. We assume that the underlying neuroanatomy, represented in these images, consists of N separate structures (or classes), c_i, $i = 1, \cdots, N$. Let subject $\mathcal{A}$ be characterized by m multi-modal images so that $\bar{I}_{\mathcal{A}}(x) \in I\!R^m$ and subject $\mathcal{B}$ be characterized by n multi-modal images so that $\bar{I}_{\mathcal{B}}(x) \in I\!R^n$, where $c_i(x)$ is the class associated spatial position $x = [x_1, x_2, x_3]^T \in \Omega_{\mathcal{A},\mathcal{B}}$, respectively. From the multi-modal images, $\bar{I}_{\mathcal{A}}(x)$ and $\bar{I}_{\mathcal{B}}(x)$, for each class c_i we jointly estimate the posterior distributions $p_{\mathcal{A}}(c_i(x))|\bar{I}_{\mathcal{A}})$ and $p_{\mathcal{B}}(c_i(x)|\bar{I}_{\mathcal{B}})$ along with the registration map, $h(x)$, that maps the space of subject $\mathcal{A}$, $\Omega_{\mathcal{A}} \subset I\!R^3$, into the space of subject $\mathcal{B}$, $\Omega_{\mathcal{B}} \subset I\!R^3$. This method is independent of the choice of m and n. Optimal inter-subject multi-modal image registration is estimated by an alternating iterative algorithm which is motivated by an expectation maximization method used in [19,12]. Our algorithm interleaves the estimation of the posteriors associated with studies $\mathcal{A}$ and $\mathcal{B}$ and the estimation of the registration map that maps $\Omega_{\mathcal{A}}$ to $\Omega_{\mathcal{B}}$.

Following [19], for each class c_i the associated data likelihood is modeled as a normal distribution with mean, μ_i, and covariance, Σ_i. That is, the probability that a voxel at spatial location x having m-dimensional intensity values $\bar{I}(x)$ belongs to class c_i, is given by

$$p(\bar{I}(x)|c_i(x), \mu_i, \Sigma_i) = \frac{1}{\sqrt{(2\pi)^N |\Sigma_i|}} e^{-\frac{1}{2}(\bar{I}(x)-\mu_i)^T \Sigma_i^{-1} (\bar{I}(x)-\mu_i)},$$

where $c_i(x)$ is the class associated with the voxel at position $x = [x_1, x_2, x_3]^T \in I\!R^3$.

Given a transformation, $h(x)$, mapping the coordinate systems $\Omega_{\mathcal{A}}$ and $\Omega_{\mathcal{B}}$ and current estimates μ_i, Σ_i, the posterior densities of the two studies can be associated with each other by using the posterior of one as the prior for the other. Using Bayes's Rule, the posterior of subject $\mathcal{A}$ becomes,

$$p_{\mathcal{A}}(c_i(x)|\bar{I}_{\mathcal{A}}) = \frac{p_{\mathcal{A}}(\bar{I}_{\mathcal{A}}(x)|c_i(x), \mu_i^{\mathcal{A}}, \Sigma_i^{\mathcal{A}}) p_{\mathcal{B}}(c_i(h(x))|\bar{I}_{\mathcal{B}})}{\sum_{k=1}^{N} p_{\mathcal{A}}(\bar{I}_{\mathcal{A}}(x)|c_k(x), \mu_k^{\mathcal{A}}, \Sigma_k^{\mathcal{A}}) p_{\mathcal{B}}(c_k(h(x))|\bar{I}_{\mathcal{B}})},$$

where $p_{\mathcal{B}}(c_i(h(x))|\bar{I}_{\mathcal{B}})$ is the transformed posterior of $\mathcal{B}$ viewed as the prior for $\mathcal{A}$. Similarly, the posterior, under the transformation $h^{-1}(x)$, of subject $\mathcal{A}$ is viewed as the prior for subject $\mathcal{B}$ is given by,

$$p_{\mathcal{B}}(c_i(x)|\bar{I}_{\mathcal{B}}) = \frac{p_{\mathcal{B}}(\bar{I}_{\mathcal{B}}(x)|c_i(x), \mu_i^{\mathcal{B}}, \Sigma_i^{\mathcal{B}}) p_{\mathcal{A}}(c_i(h^{-1}(x))|\bar{I}_{\mathcal{A}})}{\sum_{k=1}^{N} p_{\mathcal{B}}(\bar{I}_{\mathcal{B}}(x)|c_k(x), \mu_k^{\mathcal{B}}, \Sigma_k^{\mathcal{B}}) p_{\mathcal{A}}(c_k(h^{-1}(x))|\bar{I}_{\mathcal{A}})}.$$

Having defined the posteriors, the parameters $\mu_i^{\mathcal{A}}, \mu_i^{\mathcal{B}}, \Sigma_i^{\mathcal{A}}$, and $\Sigma_i^{\mathcal{B}}$ are updated via their expected values:

$$\mu_i^{\mathcal{A}} = \frac{\int_{\Omega_{\mathcal{A}}} p_{\mathcal{A}}(c_i(x)|\bar{I}_{\mathcal{A}})\bar{I}_{\mathcal{A}}(x)dx}{\int_{\Omega_{\mathcal{A}}} p_{\mathcal{A}}(c_i(x)|\bar{I}_{\mathcal{A}})dx}, \;\; \mu_i^{\mathcal{B}} = \frac{\int_{\Omega_{\mathcal{B}}} p_{\mathcal{B}}(c_i(x)|\bar{I}_{\mathcal{B}})\bar{I}_{\mathcal{B}}(x)dx}{\int_{\Omega_{\mathcal{B}}} p_{\mathcal{B}}(c_i(x)|\bar{I}_{\mathcal{B}})dx},$$

$$\Sigma_i^{\mathcal{A}} = \frac{\int_{\Omega_{\mathcal{A}}} p_{\mathcal{A}}(c_i(x)|\bar{I}_{\mathcal{A}}) \left[(\bar{I}_{\mathcal{A}}(x) - \mu_i)(\bar{I}_{\mathcal{A}}(x) - \mu_i)^T\right] dx}{\int_{\Omega_{\mathcal{A}}} p_{\mathcal{A}}(c_i(x)|\bar{I}_{\mathcal{A}})dx},$$

and

$$\Sigma_i^{\mathcal{B}} = \frac{\int_{\Omega_{\mathcal{B}}} p_{\mathcal{B}}(c_i(x)|\bar{I}_{\mathcal{B}}) \left[(\bar{I}_{\mathcal{B}}(x) - \mu_i)(\bar{I}_{\mathcal{B}}(x) - \mu_i)^T\right] dx}{\int_{\Omega_{\mathcal{B}}} p_{\mathcal{B}}(c_i(x)|\bar{I}_{\mathcal{B}})dx}.$$

Using the Kullback-Leibler divergence, $D_{KL}(\cdot,\cdot)$, the transformation is estimated via the minimization

$$\hat{h}(x) = \arg\min_{h(x)} \frac{1}{|\Omega_{\mathcal{A}}|} \int D_{KL}(p_{\mathcal{A}}(\bar{c}(x)|\bar{I}_{\mathcal{A}}), p_{\mathcal{B}}(\bar{c}(h(x))|\bar{I}_{\mathcal{B}}))dx.$$

The transformation $h(x)$ is constrained to be a diffeomorphism following the fluid formulation of [1]. The whole process is repeated until convergence.

2.1 Kullback-Leibler Divergence

Mutual information is a commonly employed similarity metric for comparing two radiometrically dissimilar images. A recent survey [13] presents a careful review of mutual information in medical image registration. This survey discusses the issues involved in defining generalized mutual information and reviews existing work in this area. The Kullback-Leibler divergence easily generalizes to any number of images.

As a distance measure between the distributions $p_{\mathcal{A}}(\bar{c}(x)|\bar{I}_{\mathcal{A}})$ and $p_{\mathcal{B}}(\bar{c}(h(x))|\bar{I}_{\mathcal{B}})$, Kullback-Leibler divergence (relative entropy), is defined as

$$\begin{aligned} &D_{KL}(p_{\mathcal{A}}(\bar{c}(x)|\bar{I}_{\mathcal{A}}), p_{\mathcal{B}}(\bar{c}(h(x))|\bar{I}_{\mathcal{B}})) = \\ &\sum_{i=1}^{N} p_{\mathcal{A}}(\bar{c}(x)|\bar{I}_{\mathcal{A}}) \log \frac{p_{\mathcal{A}}(\bar{c}(x)|\bar{I}_{\mathcal{A}})}{p_{\mathcal{B}}(\bar{c}(h(x))|\bar{I}_{\mathcal{B}})} = \\ &\qquad \mathcal{E}\left[\log \frac{p_{\mathcal{A}}(\bar{c}(x)|\bar{I}_{\mathcal{A}})}{p_{\mathcal{B}}(\bar{c}(h(x))|\bar{I}_{\mathcal{B}})}\right]. \end{aligned}$$

From an information theoretic viewpoint [4], this measure can be interpreted as the inefficiency of assuming that $p_{\mathcal{B}}(\bar{c}(h(x))|\bar{I}_{\mathcal{B}})$ is true when $p_{\mathcal{A}}(\bar{c}(x)|\bar{I}_{\mathcal{A}})$ is true. That is, if we have a model expressed as a probability density $p_{\mathcal{B}}(\bar{c}(h(x))|\bar{I}_{\mathcal{B}})$, we can then measure how far an observation, also expressed by a probability density, $p_{\mathcal{A}}(\bar{c}(x)|\bar{I}_{\mathcal{A}})$, deviates from $p_{\mathcal{B}}(\bar{c}(h(x))|\bar{I}_{\mathcal{B}})$ using Kullback-Leibler divergence. The second equality above simply highlights the notion that Kullback-Leibler divergence is the expected logarithm of the likelihood ratio of $p_{\mathcal{A}}(\bar{c}(x)|\bar{I}_{\mathcal{A}})$ to $p_{\mathcal{B}}(\bar{c}(h(x))|\bar{I}_{\mathcal{B}})$.

Kullback-Leibler divergence is non-negative, and is zero if and only if the probabilities are equal. However, Kullback-Leibler divergence is not a true distance in the strict mathematical sense since it is not symmetric. We are also investigating the use of a symmetric generalized form of the Kullback-Leibler divergence motivated by the work of Renyi [18]. Such a symmetric form would eliminate the asymmetry in the divergence computation.

2.2 Registration

In this section, we discuss the process of registering the posterior of subject $\mathcal{B}$ to the posterior of subject $\mathcal{A}$. We estimate the transformation $h(x)$ that maps the space of the prior, $\Omega_{\mathcal{A}}$, into that of the posterior, $\Omega_{\mathcal{B}}$, via the following average Kullback-Leibler minimization:

$$\hat{h}(x) = \arg\min_{h(x)} \frac{1}{|\Omega_A|} \int_{\Omega_{\mathcal{A}}} D_{KL}(p_{\mathcal{A}}(\bar{c}(x)|\bar{I}_{\mathcal{A}}), p_{\mathcal{B}}(\bar{c}(h(x))|\bar{I}_{\mathcal{B}}))dx$$

where $h(x) \in I\!R^3$ is the high-dimensional vector field describing local variation between the prior spatial coordinate system and the posterior spatial coordinate system and $D_{KL}(\cdot,\cdot)$ is the Kullback-Leibler divergence.

We apply the theory of large deformation fluid diffeomorphisms of [9,11] to require that the deformation, $h(x)$, be the solution of the Lagrangian o.d.e., $\frac{d}{dt}h(x,t) = v(h(x,t),t)$. The optimal transformation is found by estimating the velocity field following the minimization:

$$\hat{v}(\cdot,\cdot) = \arg\min_{v(\cdot)} \frac{1}{|\Omega_{\mathcal{A}}|} \int_{\Omega_{\mathcal{A}}} D_{KL}(p_{\mathcal{A}}(\bar{c}(x)|\bar{I}_{\mathcal{A}}), p_{\mathcal{B}}(\bar{c}(h(x))|\bar{I}_{\mathcal{B}}))dx + \int_0^T \int_{\Omega_{\mathcal{A}}} ||Lv(x,t)||^2 dxdt$$

where the second term is the fluid formulated regularization. Following Christensen's greedy algorithm for propagating templates [3], we compute the variation of the average Kullback-Leibler divergence term,

$$\frac{\partial}{\partial h} \frac{1}{|\Omega_{\mathcal{A}}|} \int_{\Omega_{\mathcal{A}}} D_{KL}(p_{\mathcal{A}}(\bar{c}(x)|\bar{I}_{\mathcal{A}}), p_{\mathcal{B}}(\bar{c}(h(x))|\bar{I}_{\mathcal{B}}))dx.$$

To compute this variation, we approximate $\log x$ near $x = 1$ using the Maclaurin series expansion. The point $x = 1$ is chosen for the expansion as it is the desired ratio of $\frac{p_{\mathcal{A}}(c_i(x)|\bar{I}_{\mathcal{A}})}{p_{\mathcal{B}}(c_i(h(x))|\bar{I}_{\mathcal{B}})}$ for ideal registration. The logarithm can be approximated up to the second order by $\log x \approx \frac{1}{\ln 2}(-\frac{1}{2}x^2 + 2x - \frac{3}{2})$. Thus, the second order approximation to Kullback-Leibler divergence becomes

$$D_{KL}(p_{\mathcal{A}}(\bar{c}(x)|\bar{I}_{\mathcal{A}}), p_{\mathcal{B}}(\bar{c}(h(x))|\bar{I}_{\mathcal{B}})) \approx -\frac{1}{\ln 2} \sum_{i=1}^{N} p_{\mathcal{A}}(c_i(x)|\bar{I}_{\mathcal{A}})$$

$$\cdot \left[-\frac{1}{2} \left(\frac{p_{\mathcal{B}}(c_i(h(x))|\bar{I}_{\mathcal{B}})}{p_{\mathcal{A}}(c_i(x)|\bar{I}_{\mathcal{A}})} \right)^2 + 2 \left(\frac{p_{\mathcal{B}}(c_i(h(x))|\bar{I}_{\mathcal{B}})}{p_{\mathcal{A}}(c_i(x)|\bar{I}_{\mathcal{A}})} \right) - \frac{3}{2} \right] =$$

$$-\frac{1}{\ln 2} \sum_{i=1}^{N} \left[-\frac{1}{2} \frac{p_{\mathcal{B}}(c_i(h(x))|\bar{I}_{\mathcal{B}})^2}{p_{\mathcal{A}}(c_i(x)|\bar{I}_{\mathcal{A}})} + 2p_{\mathcal{B}}(c_i(h(x))|\bar{I}_{\mathcal{B}}) - \frac{3}{2} \right].$$

Therefore, we compute the variation of the average Kullback-Leibler divergence term as follows,

$$\frac{\partial}{\partial h} \frac{1}{|\Omega_{\mathcal{A}}|} \int_{\Omega_{\mathcal{A}}} D_{KL}(p_{\mathcal{A}}(\bar{c}(x)|\bar{I}_{\mathcal{A}}), p_{\mathcal{B}}(\bar{c}(h(x))|\bar{I}_{\mathcal{B}}))dx \approx$$

$$-\frac{1}{\ln 2} \frac{1}{|\Omega_{\mathcal{A}}|} \frac{\partial}{\partial h} \int_{\Omega_{\mathcal{A}}} \sum_{i=1}^{N} \left[-\frac{1}{2} \frac{p_{\mathcal{B}}(c_i(h(x))|\bar{I}_{\mathcal{B}})^2}{p_{\mathcal{A}}(c_i(x)|\bar{I}_{\mathcal{A}})} + 2p_{\mathcal{B}}(c_i(h(x))|\bar{I}_{\mathcal{B}}) - \frac{3}{2} \right] =$$

$$\frac{1}{\ln 2} \frac{1}{|\Omega_{\mathcal{A}}|} \int_{\Omega_{\mathcal{A}}} \sum_{i=1}^{N} \left[\frac{p_{\mathcal{B}}(c_i(h(x))|\bar{I}_{\mathcal{B}})}{p_{\mathcal{A}}(c_i(x)|\bar{I}_{\mathcal{A}})} - 2 \right] \nabla p_{\mathcal{B}} \Big|_{c_i(h(x))}^{T} \cdot$$

The velocity field at each iteration is updated by solving the p.d.e.,

$$Lv(x,t) = \frac{\partial}{\partial h} \int_{\Omega_{\mathcal{A}}} D_{KL}(p_{\mathcal{A}}(\bar{c}(x)|\bar{I}_{\mathcal{A}}), p_{\mathcal{B}}(\bar{c}(h(x))|\bar{I}_{\mathcal{B}}))dx,$$

where $L = a\nabla^2 + b\nabla \cdot \nabla + cI$ is the Navier-Stokes operator.

3 Atlas-Based Image Segmentation

The above framework for multi-modal image registration can be applied to unify atlas-based image segmentation and registration. The problem of high-dimensional multi-modal image registration reduces to the problem of atlas-based image segmentation where subject $\mathcal{B}$, in Section 2, is replaced by an atlas prior with known probability densities.

Given a multi-modal image data set, $\bar{I}(x)$, and a geometric prior $p(c_i(x))$, image segmentation lends itself to the Bayesian paradigm in which inference about a class, c_i, of each of the voxels is based on the posterior distribution $p(c_i(x)|\bar{I})$ given by Bayes's Rule. To apply Bayes's Rule, the prior and likelihood need to be in the same coordinate system, and hence, a registration map is required. If the prior and likelihood are not in the same coordinate system then the estimation of the posterior is subject to overlap error due to spatial misalignment. Working within the probabilistic framework defined in Section 2 unifies the approach of estimating both the posterior and the registration map that matches the coordinate spaces of the prior and posterior.

4 Results

To evaluate the performance of the algorithm we define a geometric prior, a known transformation, and two synthetic images whose radiometric characteristics are statistically similar to actual T1- and T2-weighted MR images. A four

class atlas prior comprising of concentric ellipses was generated using Matlab. A composite label image is given by the superposition of the individual classes. Both the atlas and composite label image are shown in Figure 1. Additionally, a transformation, $h(x)$, was constructed using sinusoidal displacements which was then applied to the composite label image. The deformed label image is shown in the left panel of Figure 2. The two synthetic images were simulated by sampling from a multi-variate Gaussian distribution with different means and covariances for each of the classes in the deformed label image.

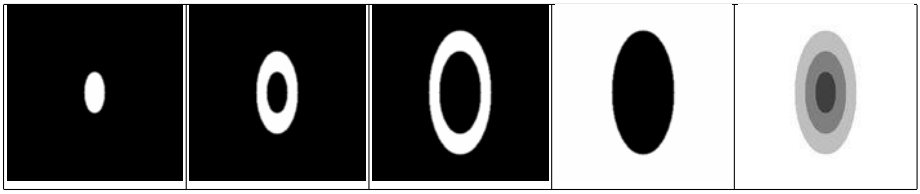


Fig. 1. Geometric Atlas Prior. Manually generated geometric four class atlas prior (first four images) and the corresponding composite labeled image.

The algorithm was was run for fifty iterations with ten steps of the high-dimensional registration per iteration. The final segmentation and deformation estimates are also shown in Figure 2.

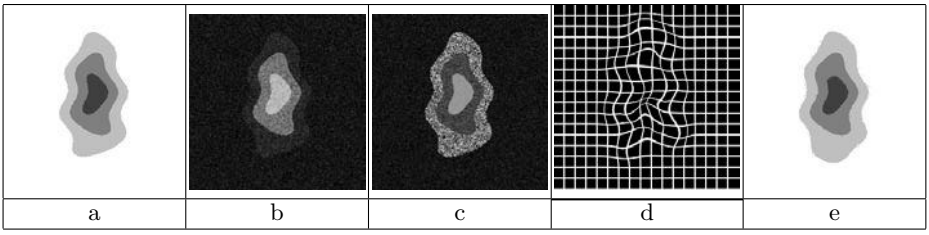


Fig. 2. Synthetic Images and Estimated deformation. Deformed label image (a), synthetic image derived from T1 samples (b), and synthetic image derived from T2 samples (c). Estimated deformation (d) and segmentation (e) following fifty iterations of the algorithm.

In order to evaluated the success of the high-dimensional registration the algorithm is run, again with fifty iterations, holding the transformation fixed to the identity map. With the transformation fixed to the identity map, the expectation maximization provides the optimal maximum likelihood solution. The final estimated segmentation is then compared to the one based on registration, with results shown in Figure 3. By examining the regions where the two segmentations differ from the ground truth label image it is clear the registration has improved the segmentation.

In both invocations of the algorithm, the class means and covariances were collected and compared. Table 1 shows the final relative norms for the estimated and actual means and covariances at the final iteration. For all classes

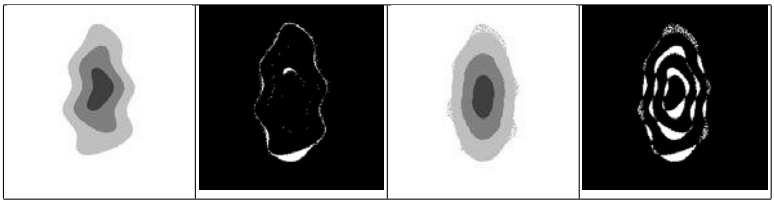


Fig. 3. Final Segmentation. The image in the upper left is the ground truth label image. The second column shows the final segmentation estimation using registration and the regions where this segmentation differs from the ground truth. The right column shows the same where the transformation is fixed to the identity map.

Table 1. Relative Norm Statistics. The first two columns of numbers are the relative means using registration (left) and fixed identity map (right) as the final iteration. The last two columns show the same for the relative covariances.

Class	$\frac{\|\|\hat{\mu}_{c_i}-\mu_{c_i}\|\|}{\|\|\mu_{c_i}\|\|}$ w/ reg.	$\frac{\|\|\hat{\mu}_{c_i}-\mu_{c_i}\|\|}{\|\|\mu_{c_i}\|\|}$ w/o reg.	$\frac{\|\|\hat{\Sigma}_{c_i}-\Sigma_{c_i}\|\|}{\|\|\Sigma_{c_i}\|\|}$ w/ reg.	$\frac{\|\|\hat{\Sigma}_{c_i}-\Sigma_{c_i}\|\|}{\|\|\Sigma_{c_i}\|\|}$ w/o reg.
c_1	0.0047	0.1216	0.0450	7.0588
c_2	0.0152	0.1168	0.2104	1.8164
c_3	0.0960	0.0939	0.0206	0.0811
c_4	0.0046	0.0046	0.0081	0.0266

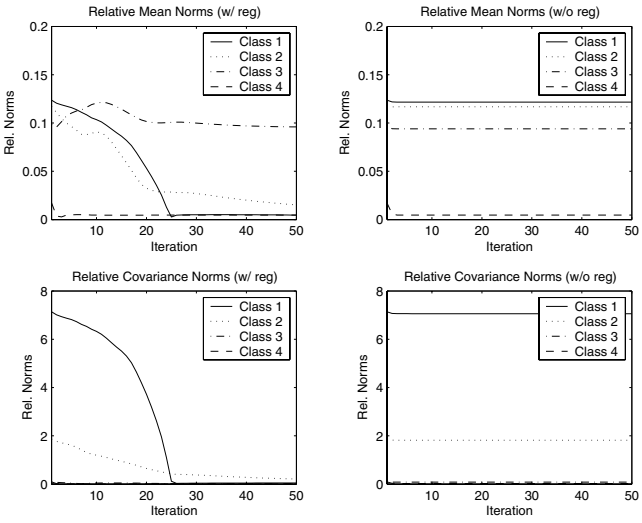
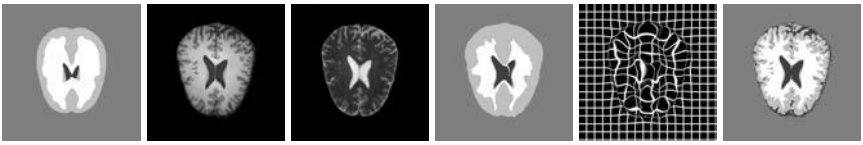


Fig. 4. Convergence. The left column shows the convergence of means and covariances using registration. The right column shows the same using the fixed identity transformation.

the registration has improved estimates for both the means and covariances. The convergence of the means and covariance estimates using registration is shown in Figure 4. From this figure we see that the estimation of the means and covari-



a b c d e f

Fig. 5. Segmentation of Normal Brain Tissue: the first three images show the hard labeling of the fuzzy atlas (a) and T1 and T2-weighted images of a normal subject (b+c). The second three images shows the results of the segmentation process: the hard labeling of the deform atlas (d), a regular grid under this deformation (e), and the final segmentation (f).

ances have converged quickly when the transformation is fixed to the identity map. When registration is added the estimates of the means and covariances continue to improve as estimation of the transformation between atlas and the subject converges. This exemplifies the effectiveness of the alternating nature of the algorithm.

From these results we see that the registration improves the segmentation by accommodating local variability. Figure 5 shows the results of this registration between a fuzzy atlas and two MR images of a normal subject. The multi-modal inter-subject image registration results are currently being developed and will be available by the time of the workshop.

Acknowledgments

The authors would like to thank Dr. Guido Gerig, of the department of computer science, Dr. Elizabeth Bullitt, of the department of surgery, and Marcel Prastawa, of the department of computer science for their insightful suggestions and assistance during the development of this manuscript. This work was supported by NIBIB-NIH grant R01 EB000219.

References

1. Gary Christensen, R Rabbit, and Michael Miller. Deformable templates using large deformation kinematics. *Transactions on Image Processing*, 5(10):1435–1447, 1996.
2. G.E. Christensen, S.C. Joshi, and M.I. Miller. Deformable templates using large deformation kniematics. *IEEE Transactions on Medical Imaging*, 16:864–877, 1997.
3. G. E. Christensen, R. D. Rabbitt, and M. I. Miller. Deformable templates using large deformation kinematics. *IEEE Transactions on Image Processing*, 5(10):1435–1447, October 1996.
4. Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory.* Wiley-Interscience, New York, 1991.
5. JG Csernansky, S Joshi, LE Wang, J Haller, M Gado, MI Miller, and U Grenander. Hippocampal morphometry in schizophrenia via high dimensional brain map ping. *Proc. Natl. Acad. Sci. USA*, 95:11406–11411, September 1998.

6. Tom Gaens, Frederik Maes, Dirk Vandermeulen, and Paul Suetens. Non-rigid multimodal image registration using mutual information. *Proceedings of Medical Image Computing and Computer-Assisted Intervention*, pages 1099–1106, 1998.
7. J. C. Gee and R. K. Bajcsy. Elastic matching: Continuum mechanical and probabilistic analysis. In A. W. Toga, editor, *Brain Warping*. Academic Press, 1999.
8. A. Guimond, A. Roche, N. Ayache, and J. Meunier. Multimodal brain warping using the demons algorithm and adaptative intensity corrections. *IEEE Transaction on Medical Imaging*, 20(1):58–69, 2001.
9. S. Joshi and M. I. Miller. Landmark matching via large deformation diffeomorphisms. *IEEE Transactions on Image Processing*, 9(8):1357–1370, August 2000.
10. Frederik Maes, Andre Collignon, and at al. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging (TMI)*, 16:187–198, April 1997.
11. Michael Miller, Sarang Joshi, and Gary Christensen. Large deformation fluid diffeomorphisms for landmark and image matching. In *Brain Warping*, pages 115–131. Wiley-Interscience, 1999.
12. Nathan Moon, E Bullitt, K van Leemput, and Guido Gerig. Automatic brain and tumor segmentation. *Medical Image Computing and Computer–Assisted Intervention (MICCAI)*, LNCS–2489:372–379, 00–00 September 2002.
13. Josien Pluim, Antoine Maintz, and Max Viergever. Mutual information based registration of medical images: a survey. *To appear in IEEE Transactions on Medical Imaging (TMI): www.isi.uu.nl/People/Josien/Papers/Pluim_TMI_2003.pdf*, 2003.
14. Kilian Pohl, William Wells, Alexandre Guimond, Kiyoto Kasai, Martha Shenton, Ron Kikinis, Eric Grimson, and Simon Warfield. Incorporating non-rigid registration into expectation maximization algorithm to segment mr images. *Medical Image Computing and Computer–Assisted Intervention (MICCAI)*, LNCS–0000:564–571, 00–00 September 2002.
15. D. Rueckert, C. Hayes, C. Studholme, P. Summers, M. Leach, and D. J. Hawkes. Non-rigid registration of breast mr images using mutual information. *Proceedings of Medical Image Computing and Computer-Assisted Intervention*, pages 1144–1152, 1998.
16. C. Studholme, D. L. G. Hill, and D. J. Hawkes. An overlap invariant entropy measure of 3d medical image alignment. *Pattern Recognition*, pages 71–86, December 1998.
17. Arthur Toga. *Brain Warping*. Wiley-Interscience, New York, 1999.
18. P. Turan. On measures of entropy and information. In *Selected Papers of Alfred Renyi*, pages 565–579. Akademiai Kiado, Budapest, 1976.
19. K. v. Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automated model-based tissue classification of mr images of the brain. *IEEE Transactions on Medical Imaging (TMI)*, 18:897–908, 1999.
20. J. West, J. Fitzpatrick, M. Wang, B. Dawant, C. Maurer, R. Kessler, and R. Maciunas. Comparison and evaluation of retrospective intermodality image registration techniques. *Proceedings of the SPIE Conference on Medical Imaging*, 1996.

A New Technique for Multi-modal 3D Image Registration

Gjenna Stippel[1], Jim Ellsmere[2], Simon K. Warfield[2], William (Sandy) M. Wells[2], and Wilfried Philips[1]

[1] Ghent University, Department TELIN, Gent, Belgium
[2] Harvard Medical School, Brigham & Women's Hospital, Boston, MA USA
http://www.gjenna.com

Abstract. In this paper we address the problem of multi-modal co-registration of medical 3D images. Several techniques for the rigid registration of multi-modal images have been developed; in one of those the Kullback-Leibler distance is used to align 2D-3D angiographic images [1]. In this paper we investigate the performance of this technique on the registration of pairs of 3D CT/US images. We study the effects of various image perturbations on the performance of the registration, and obtain promising results.

1 Introduction

Registering intraoperative ultrasound (US) to preoperative images like computed tomography (CT) or magnetic resonance (MR) remains a challenging problem in the field of image guided surgery. For many surgical applications, ultrasound provides adequate clinical information to carry out the required intervention. There are some applications, however, where being able to interpret the ultrasound images in the context of the higher quality preoperative imaging has been shown to be helpful [2].

In this paper, we introduce a new rigid registration technique that aligns US and CT images by applying linear transformations to the coordinate system in order to minimize the Kullback-Leibler distance (KLD) between the observed joint intensity distribution, and a reference distribution representing the prior knowledge of the expected joint intensity histogram when the images are aligned properly [1].

We calibrate our KLD registration by taking an ultrasound scan, and manually aligning it with a CT scan to establish the fixed reference distribution. Afterwards, as more 3D ultrasound acquisitions are acquired over the course of the surgery, this calibration does not need to be repeated. Hence, the cost of the calibration is amortized over all the follow up scans during the procedure. If there are many scans, the calibration cost is a small percentage of the total effort.

The performance of the method and its robustness under various image perturbations is studied.

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 244–253, 2003.

2 The Registration Technique

The proposed technique is an intensity based registration algorithm. As said, in a first calibration step we align one image pair consisting of an image of the same organ from each modality manually. From this initial manual alignment a joint intensity histogram $P(g_1, g_2)$ with $0 \leq g_1, g_2 \leq 255$ is constructed as follows:

Let I_1 and I_2 be the intensity values of the two images, and let X_1 and X_2 be their image domains respectively. Since the images are aligned, we have $X_1 = X_2$. Let $J := \{(x, i_1(x), i_2(x)) | x \in X_1\},\ (i_1 \in I_1,\ i_2 \in I_2)$. Now the joint intensity histogram P is the (two-dimensional) histogram of J, i.e., $P(g_1, g_2) := |\{(a, b, c) \in J | b = g_1 \wedge c = g_2\}|\ (0 \leq g_1, g_2 \leq 255)$, where $|S|$ denotes the cardinality of the set S.

This histogram serves as a reference distribution in all further alignments. To register the new image automatically, we can iteratively translate and rotate one of the two images (the "moving image") while keeping the other image fixed (the "fixed image"), thereby aiming to determine the rigid transformation for which the resulting joint intensity distribution corresponds most with our reference distribution. The measure we use to quantify the similarity of the two histograms is the so-called Kullback -Leibler distance.

2.1 Kullback-Leibler Distance

Let the reference distribution be denoted by P_{ref} and the observed joint intensity distribution after the transformation T has been applied to the moving image by P_T. Now the Kullback-Leibler distance between the two distributions is defined as follows:

$$D(P_T || P_{\text{ref}}) := \sum_{i_1, i_2 = 0}^{255} P_T(i_1, i_2) \log \frac{P_T(i_1, i_2)}{P_{\text{ref}}(i_1, i_2)}.$$

The Kullback-Leibler distance is not a distance in the mathematical sense, since in general it is not symmetric: $D(P_{\text{ref}} || P_T) \neq D(P_T || P_{\text{ref}})$ and thus does not satisfy one of the three criteria any distance has to satisfy. Two other distance properties are met and are the most important for our purpose [3,4]:

1. $D(P_T || P_{\text{ref}}) \geq 0$,
2. $D(P_T || P_{\text{ref}}) = 0$ if and only if $P_T = P_{\text{ref}}$

These two properties make the Kullback-Leibler distance a useful measure to determine how far a particular alignment of the two images is from our "reference alignment" obtained from the manual calibration. For an image pair that is identical to the image pair used in the manual calibration, we observe that when these two images are perfectly aligned, the joint intensity histogram coincides exactly with the reference histogram, so in that case the KLD equals zero; otherwise the KLD is strictly positive.

The idea behind the registration technique is thus, to find a transformation T_0, acting on the moving image, that minimizes the KLD between the joint intensity distribution P_{T_0} and the reference distribution P_{ref}. Or, in formula:

$$T_0 = \arg\min_T D(P_T||P_{\text{ref}}).$$

In this paper we restrict ourselves to translations only because of their computational simplicity. The way to deal with the problem when we include rotations also is exactly the same; we only need to conduct our search in a six-dimensional space then, instead of a three-dimensional one.

Moreover, registration of volumetric images using translations exclusively has practical applications: if the images are taken under approximately the same angle with respect to the skin surface and the imaged organs do not rotate with respect to the skin surface themselves either (e.g., as with the brain), rotations have no significant impact on the registration. So in this case it is sufficient to use translations only.

The strategy we follow in the technique to find the optimal alignment (i.e., the one closest to the reference alignment) is as follows:

We keep the CT scan fixed during the entire procedure. The US image can be translated in any direction. Using a steepest descent optimization method we search, by iterative translations, the position for the moving image at which the KLD between the corresponding joint intensity distribution and the reference distribution is minimal.

Several questions arise. How robust is the method against "image variability" with respect to:

- geometric deformations; the brain is fairly rigid but it still differs in shape somewhat from one scan to another.
- overall brightness; due to different machine settings or anatomical differences, like the amount of hair an infants has on his head, the US images can differ in overall brightness from one scan to another.
- noise; even the slightest difference of the head of the transducer influences the speckle noise in the US image.

All these questions will be addressed in Section 6.

KLD has been previously used by Chung et al. [1] to do 2D-3D registration of angiographic image data.

3 Dataset

To generate our test set, we used a 3D abdominal phantom (CIRS, Norwalk, VA). The CT scan (512x512, 1mm slices) was acquired using the Somatoform Plus4 (Siemens Medical Systems, Iselin, NJ). Two 3D US images of the liver (left lobe, right lower lobe) were generated using the Stradx software (Cambridge University, Cambridge, UK). The 3D US system consists of a Lynx ultrasound unit (BK Medical Systems, Wilmington, MA) and a miniBIRD tracking system (Ascension Technology, Burlington, VT). For the calibration the CT and US images were manually aligned using the Slicer software (Brigham & Women's Hospital, Boston, MA). In Figure 1 two 2D sections that are aligned, one of the CT scan and one of the US scan, are shown. The size of one 3D US image is: 460x324x18 voxels.

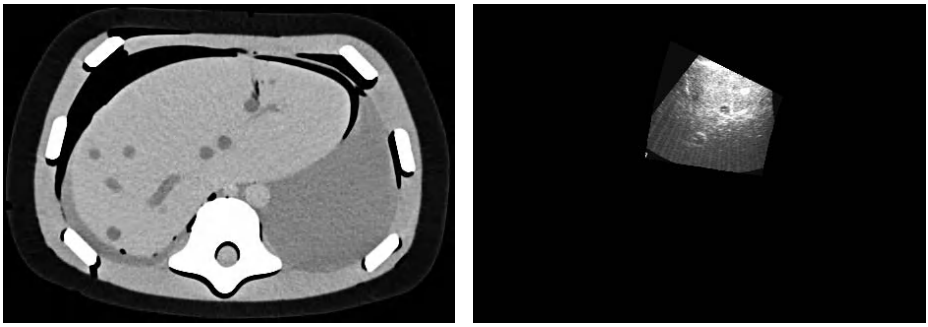


Fig. 1. 2D section of volumetric CT and US scans as used in the experiments.

4 Image Perturbations

In order to study the effects of the aspects mentioned in Section 2.1 on the proposed technique, we have perturbed the US images in various ways:

Geometrical distortion:

- we rescaled the image to 90 % of the original size in the x-direction

Brightness:

- we applied gamma-correction (with $\gamma = 0.5$) on the image to make it darker

Noise:

- we applied a Lee filter to the image [5], using a 7×7 window, and an estimated-noise-variance parameter of 250
- we applied a median filter to the image, using a 5×1 window. Visually this choice of the window size yields the best results regarding speckle suppression without blurring the image significantly
- we added multiplicative (speckle) noise to the image artificially. If we call the image I, we added the noise using the equation $J = I + n * I$, where n is uniformly distributed random noise with mean 0 and variance 0.1 or 0.3

With all these images we performed the experiment as explained in Section 5.

5 Experiment

The experiment we performed was as follows: from the manual alignment of the US image of the left lobe of the liver and the CT scan (see Section 3) we constructed our reference joint intensity histogram. We keep the CT scan fixed during the entire procedure. The US image is moved from 100 pixels to the left to 100 pixels to the right in the x-, y- and z-direction with steps of 4 pixels. At each position a new joint intensity histogram is constructed, and the KLD between this joint intens ity histogram and the reference histogram is calculated.

In Figure 2 the resulting graphs of the KLD between the unprocessed US image of the left lobe and the CT scan are displayed, when translated in the x-, y- and z-direction. In the left subfigure the translations have been performed with steps of 4 pixels, in the right subfigure with steps of 1 pixel. This demonstrates the typical shape of a KLD graph, with its minimum at the correct position.

Our ultimate goal in this work is to register more US images that are taken during the course of the surgery with the CT scan using the same reference distribution acquired from the calibration. Since, obviously, these will not be exactly the same, we repeated the experiment described above with the artificially perturbed US images (as explained in Section 4). We translate the perturbed US images with respect to the original CT scan, construct a new joint intensity histogram in every step again, and calculate the KLD between this joint intensity histogram and our standard reference distribution. In these cases we only do the translations in the x-direction from 25 pixels to the left to 25 pixels to the right, with steps of 1 pixel.

We wish to see the influence of these perturbations on the performance of the registration with respect to:

1. Accuracy
2. Speed of convergence
3. Robustness

As a measure of the accuracy we take the position of the minimum of the KLD graph. Since none of the perturbations we perform involves any translation of the image, the optimal alignment is at exactly the same position as for the original image (i.e., at 0 pixels translation). So, any number of pixels the minimum of the KLD graph corresponding to a perturbed image is next to this position, can be considered a misalignment.

The speed of convergence and the robustness are estimated by the "pointedness" of the graph. We see in Figures 3 and 4 that in general, the shape of the KLD graphs is as follows: there is one global minimum, and in the direct neighbourhood of that the graph has a "V"-shape. With the "pointedness" of the KLD graph we mean the acuteness of the angle of this "V". As a way of measuring it , we measure the width of this "V" at various heights:

If m is the x-value at which the KLD graph is minimal, then we define $b := \frac{\mathrm{kld}(m-1)+\mathrm{kld}(m+1)}{2}$ as the "base point". We define the base point like this, because the value of the KLD graph at the global minimum can sometimes be significantly lower than the values at the two points adjacent to it.

We then measure the width of the KLD graph at the heights: $b,\ b+0.001,\ b+0.002, \ldots,\ b+0.008$. If $t > 0$, then the width w_t at height t is:

$$w_t := \max\{x|\mathrm{kld}(x) \leq t\} - \min\{x|\mathrm{kld}(x) \leq t\}.$$

In Figure 4 we plotted $w_b,\ w_{b+0.001}, \ldots,\ w_{b+0.008}$ for the KLD graphs of all perturbed US images.

6 Results

We performed the experiments as explained in the text. Figure 2 consists of two subimages. In the left subimage the KLD graphs are displayed for when we move the images in the x- and the y-direction from 96 pixels to the left of the optimal alignment to 96 pixels to the right with steps of 4 pixels. In the right subimage the same KLD graph for the translations from 25 pixels to the left of the optimal alignment to 25 pixels to the right (with steps of 1 pixel) and in the z-direction fr om -18 to +18 pixels of the optimal alignment are displayed. These figures illustrate what (a one dimensional intersection of) the KLD graphs usually look like.

Figures 3 and 4 show us the KLD graphs of the various preprocessed images. In these cases, we only move the images in the x-direction.

The first thing we can conclude from these graphs is that the translation value (x-value) at which the KLD reaches its minimum remains 0 in all cases. (In our experiments 0 is the correct alignment). This demonstrates that with regard to finding the correct alignment the proposed technique is robust against image variability in small geometrical deformations, brightness and noise level.

In Figure 4 we compare the pointedness of the KLD graphs, as explained in Section 5. In general a more pointed KLD graph, i.e. a graph in which the angle that the minimum and the surrounding function values make is more acute, is preferable. Since in this case the KLD function values approach the global minimum more rapidly if the x-value approaches 0, this minimum is approached quicker with a steepest descent method. Furthermore, if the KLD graph is less pointed, i.e. the shap e of the graph is more shallow, disturbances by noise are likelier to create local minima. Local minima are undesirable since they can attract the optimizer (e.g. steepest descent method) to the wrong optimum, thereby failing to align the images properly.

In the graphs in Figure 4 the shape of the KLD graphs reveals itself in the slope of the corresponding curves: when the KLD graph is shallow, the corresponding curve in Figure 4 is very steep, since the width of the KLD graph increases fast with the height. A more pointed KLD graph results in a less steep curve in Figure 4.

We can see from the graphs in the right subfigure in Figure 4 that a small geometrical deformation hardly has any influence on the registration performance. The corresponding graphs remains very close to that of the original image. Median filtering does not have much effect either.

Furthermore after applying the gamma correction, and making the image darker, the KLD graphs becomes more pointed, thereby indicating a faster registration. This is not unexpected: by lowering the brightness of the image a lot of speckle noise is removed while mainly the important structures remain visible.

Lee filtering the US image and adding speckle noise results in slightly more pointed KLD graphs.

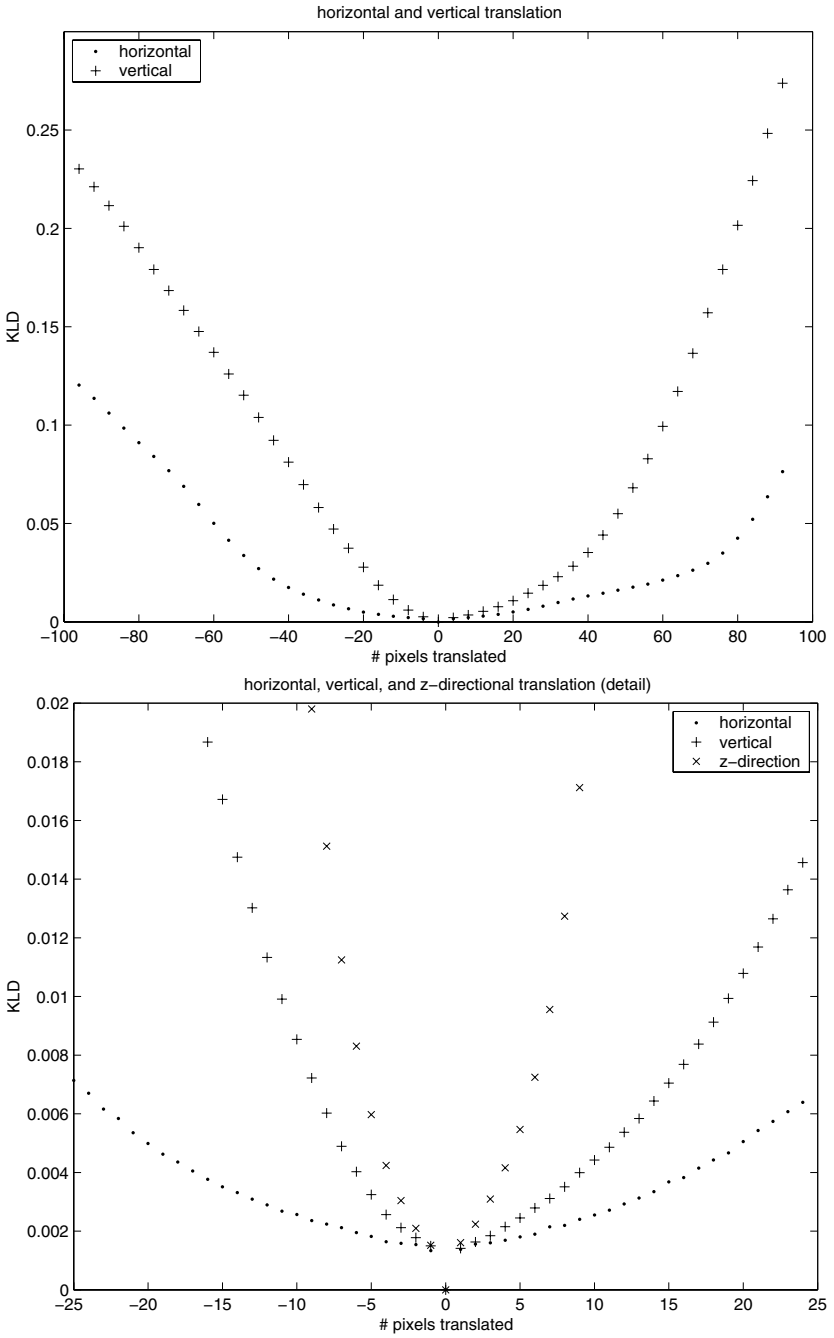


Fig. 2. KLD graphs of translations of the unprocessed US image of the left lobe.

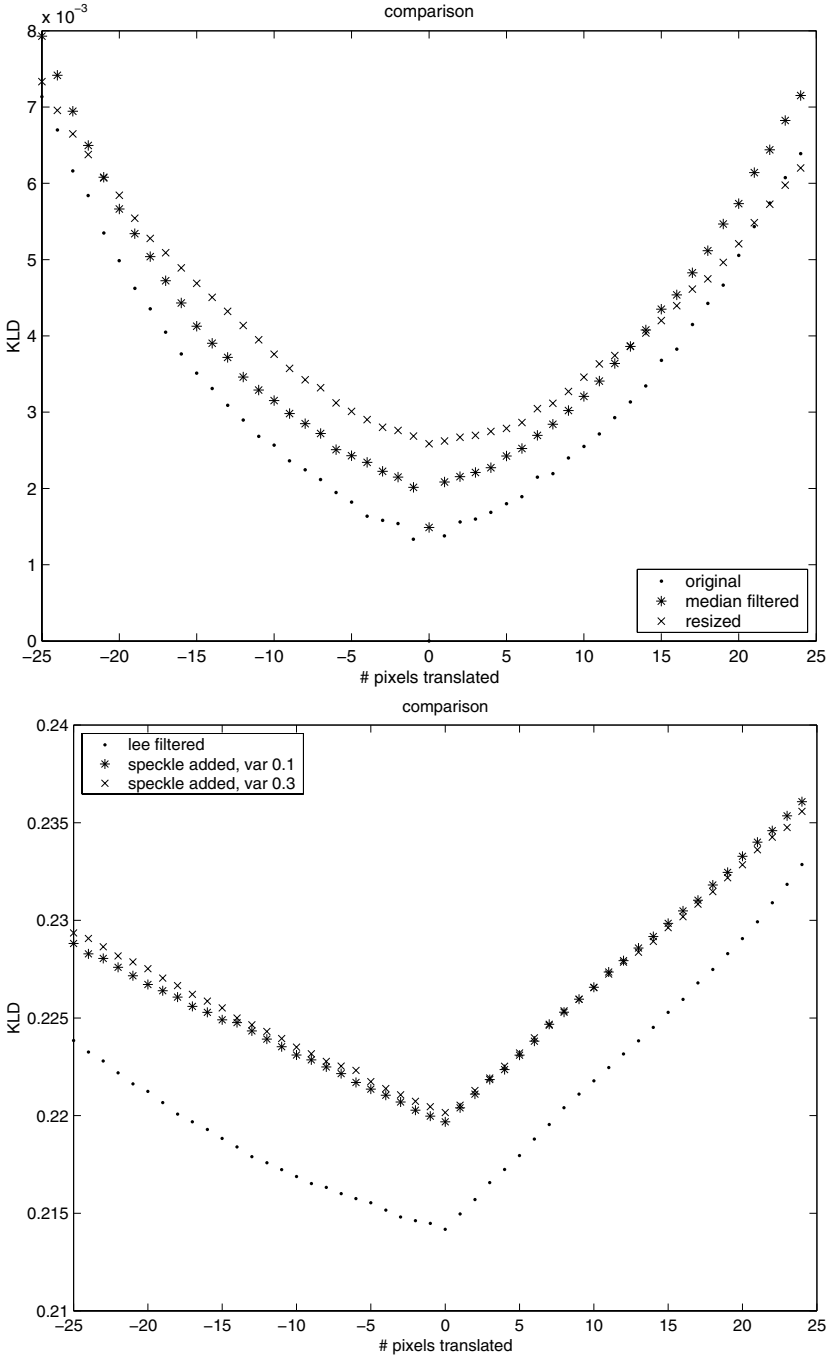


Fig. 3. KLD graphs of the perturbed images of the US image of the left lobe.

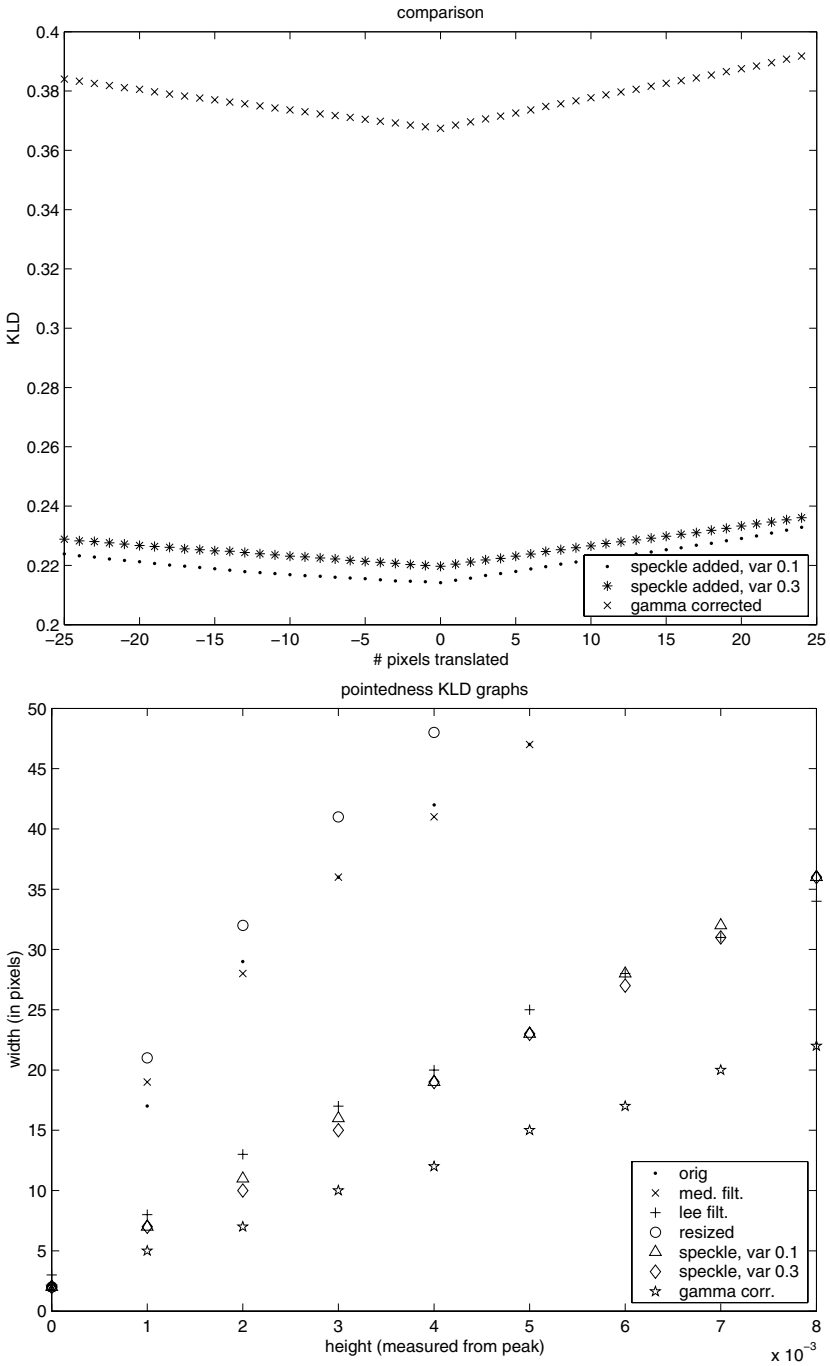


Fig. 4. Top: KLD graphs of the perturbed images of the US image of the left lobe **Bottom:** Pointedness of KLD graphs.

7 Conclusion

In this paper we introduced a novel technique for the rigid registration of volumetric multi-modal medical images. We investigated its performance on a test set of a pair of CT and volumetric US images of an abdomen phantom. The proposed technique successfully succeeds in aligning the US-CT pair, also after the various perturbations of the US image. We are currently conducting further tests on human images to assess its suitability for clinical application.

Acknowledgements

This research was supported by the following grants: NSF ERC grant under Johns Hopkins Agreement 8810274, NIH grants P41RR13218, P01CA67167.

References

1. Albert C.S. Chung, William M. Wells III, Alexander Norbash, and W. Eric L. Grimson, "Multi-modal image registration by minimising kullback-leibler distance," in *Proceedings of MICCAI*, 2002, pp. 564–571.
2. J. Ellsmere, J.A. Stoll, D.W. Rattner, D. Brooks, R. Kane, W.M. Wells III, R. Kikinis, and K.G. Vosburgh, "Integrating preoperative ct data with laparosopic ultrasound images facilitates interpretation.," *Society of American Gasrointestinal Endoscopc Surgeons*, vol. 17, pp. S296, 2003.
3. T.M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., 1991.
4. S. Kullback, *Information Theory and Statistics*, Dover Publications, Inc., 1968.
5. J. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, no. 2, pp. 165–168, 1980.

Multimodal Registration Using Patch Algorithm

P. Zhilkin, M.E. Alexander, and C.D. Mansfield

National Research Council of Canada, Institute for Biodiagnostics
435 Ellice Ave, Winnipeg, MB R3B 1Y6 Canada

Abstract. Images of different modality are registered using a unimodal image registration program ("Patch Algorithm"). An affine model for the displacement field is considered. Prior to registration the images are convolved with a set of Gabor filter (quadrature) pairs tuned to certain orientations and scales, and a gradient of local phase (local frequency) is computed for the filtered images. Local frequency representation of an image is relatively insensitive to changes in illumination conditions, and may enable common features between images of different modality to be captured. To demonstrate the algorithm, a series of 2-dimensional T_2-weighted magnetic resonance images of brain was registered against a T_1-weighted reference image. Also, registration of a near infrared hyperspectral datacube was shown to improve recovery of localized skin chromophore features for facial imaging, demonstrating its potential to obtain high spatial resolution diagnostic information.

1 Introduction

The correct alignment of images is a vital preprocessing step to ensure accurate and reliable diagnosis when analyzing multiple medical images. There is a growing need for automatic alignment of images arising from several modalities – for example, near infrared (NIR) images from different spectral bands; or T_1- and T_2-weighted magnetic resonance (MR) images. In all cases, the relation between intensities of corresponding features in the pair of images is unknown. A popular class of methods uses the mutual information [1] as a measure of alignment between a pair of multimodal images, since this method does not rely on prior knowledge of the intensity relations. Some years ago, Fleet and Jepson [2] proposed using local phase information to compute optical flow in a series of images. Liu et al. [3] proposed a similar method for registration of image pairs using local (spatial) frequency in the images. Their method makes use of the property that – unlike the local intensity information – local frequency is invariant to changes in contrast. Thus, local phase-based methods are also suitable for multimodal image registration. In this paper we describe a method that extracts local phase using quadrature Gabor filters tuned to different widths and orientations, providing the preprocessing stage for applying an efficient unimodal "patch" registration algorithm that we have earlier developed [4,5].

2 Method

The Patch Algorithm (PA) [4,5] was developed for registering images of same modality. It employs a linearized image matching equation

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 254–260, 2003.

$$\Delta I \equiv I' - I \approx eI + \mathbf{u} \cdot \nabla I \tag{1}$$

where I' is the reference image, I the target image, $\mathbf{u}$ a small displacement field to be determined, and e the mean intensity change between the two images. The unknown displacement field is usually described by an affine or nonlinear B-spline model [6]. To extend the validity of this equation to larger displacements, the images are convolved with band-pass filters implemented at decreasing spatial scales of resolution. This has the effect of removing fine-scale image features and thus of avoiding false matches between features. Equation (1) is integrated over a patch (usually a rectangle in 2D, and a prism in 3D) in the image domain, yielding a linear equation with respect to the unknown parameters defining the displacement field $\mathbf{u}$. One can cast registration as a linear regression problem, and by selecting a sufficient number of patches and filters (each patch/filter combination providing an equation for the regression), an accurate estimate of these parameters may be obtained. After the image has been registered at a coarse resolution, the disparity between corresponding image features is reduced, and the image may then be registered at successively finer resolutions. This multiresolution approach allows the registration of images with relatively large displacement (the magnitude of which determines how coarse a level of resolution is required to successfully align the images), and reduces chances of getting a false match. Integration over a patch (together with band-pass filtering) also renders the algorithm robustness to noise.

In order for the PA to register images of different modality we first compute the magnitude of the local phase gradient (local frequency (LF)) [7] in each image using Gabor quadrature filter pairs. We incorporate these filters as a pre-processing stage for the PA. The Gabor filters, each with a certain scale σ and tuned to a particular orientation (U,V) are generated from

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \exp(2\pi j(Ux + Vy)) \tag{2}$$

Images are convolved with real and imaginary parts of the Gabor filter, and gradient of local phase is computed. The LF representations of the images are computed by summing up the magnitude-squared gradients for each Gabor filter used. Then the LF images are submitted to the PA for unimodal registration. The registered image is obtained by transforming the target image according to the affine transformation found by the PA.

If the 'true' transformation is known, we may estimate the quality of registration by computing the following geometric warping index (WI) [8] which for the affine transformation $\mathbf{u}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$, becomes

$$WI = \frac{1}{N} \sum_k \left\| \left(\mathbf{A}^{-1} - \mathbf{A}_\mathbf{r}\right) \cdot \mathbf{x}_\mathbf{k} + \left(-\mathbf{A}^{-1} \cdot \mathbf{b} - \mathbf{b}_\mathbf{r}\right) \right\| \tag{3}$$

where $\{\mathbf{A},\mathbf{b}\}$ denote the matrix and vector representing the known 'true' transformation parameters, and $\{\mathbf{A}_\mathbf{r},\mathbf{b}_\mathbf{r}\}$ those restored by registration. Index k ranges over all N pixels of the images.

3 Results and Discussion

The PA enhanced for multimodal registration was tested on an artificially generated series of magnetic resonance images (for which the 'true' affine registration parameters are known, and therefore the WI, equation (3), may be computed), and a real NIR hyperspectral datacube, each slice of which effectively represents a separate modality for the purpose of registration. Linear interpolation was used for image reconstruction throughout. The orientation and scale of Gabor filters used to compute LF images were determined experimentally by visual inspection of the images obtained prior to registration, and checking how well the PA can register such images.

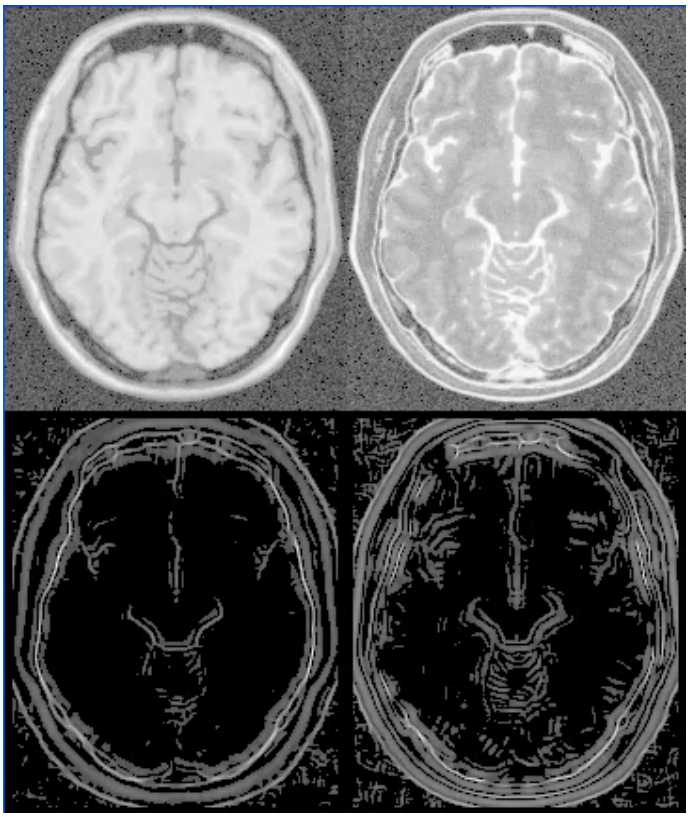


Fig. 1. Row 1: T_1 (left) and T_2 (right) MR images. Row 2: LF representation of the T_1 image (left) and a typical LF representation of a T_2 image (right).

Series of 2-Dimensional MR Images

T_1- and T_2-weighted MR images (217x181x181 1mm cubic voxels) were obtained from [9,10]. A slice was selected from both images. The T_1 slice was used as a refer-

ence image, and T_2 slice as a basis for generating a series of randomly misaligned images to register against the reference image (row 1 in Fig. 1). A unit matrix and zero translation vector were randomly perturbed, and the resulting affine transformations were applied to the T_2 image using sinc interpolation. 50 images were generated, with median displacement of approximately 2.8 pixels, maximal displacement up to 8-9 pixels. Two Gabor filters were used to compute LF representation of the images. For the T_1 reference image the filters were tuned to (U=0.09, V=0.00, σ=2.0) and (U=0.00, V=0.09, σ=2.0); for all T_2 images the filters were tuned to (U=0.11, V=0.00, σ=2.0) and (U=0.00, V=0.11, σ=2.0). In both cases the filter masks were computed using equation (2) with origin located at the center of the mask. The size of the real filter was 5x5, imaginary 9x9. Visual inspection of the LF images (that look similar to the images in row 2 in Fig. 1) indicates that such filters provide a sufficient number of details for most of the images in the series, and such images may be suitable for unimodal registration. Nevertheless, it should be mentioned that the LF images come from different images, and intensity based registration methods may be unable to register such images. Although for several generated T_2 images their LF images appear different and thus seem less suitable for unimodal registration, they were used as they are. Further individual tuning of the Gabor filters for these images may improve the situation, but that was not a part of this study.

The LF representations of T_2 images were registered against the LF representation of the T_1 image. The derived affine transformations were applied to the randomly misaligned T_2 images for realignment. The WI, as a measure of registration accuracy, was computed for each image pair. It was found that (for the series) the median WI of 2.77 prior to registration was reduced to 0.91 after registration. Visual inspection of the registered images indicated that the registered T_2 images have a more accurate match with the T_1 image than the unregistered T_2 images. In general, the misalignment was reduced in most of the image pairs. As an example, the unregistered and registered T_2 images superimposed with the edges of the reference T_1 image are presented in Fig. 2.

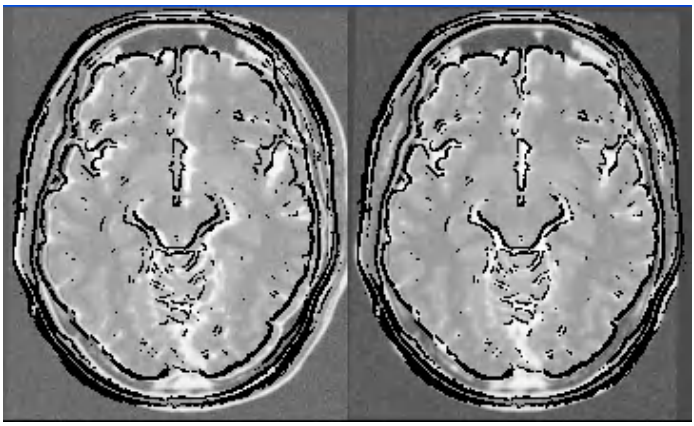


Fig. 2. T_2 MR image superimposed with edges of T_1 image before registration (left) and after registration (right).

Near Infrared Hyperspectral Datacube

NIR hyperspectral imaging is a medical imaging technique that provides non-invasive, in-vivo, diagnostic information relating to the distribution of skin chromophores. A large number of images are recorded at sequential wavelengths creating a hyperspectral datacube that consists of two spatial dimensions and a third spectral dimension. Thus, tracking each pixel along the datacube's wavelength axis will yield a low-resolution spectrum, e.g. a 256 x 256 array contains 65536 individual NIR spectra. Each of these spectrum encompasses a number of relevant biological spectroscopic features, e.g. oxygenated and deoxygenated hemoglobin and water, which upon relevant spectroscopic analysis can generate single images of relative oxygenated and deoxygenated hemoglobin concentration, blood volume, oxygen saturation and hydration. The distributions of these parameters reflect physiological changes occurring in the tissue that provides diagnostic information on its condition, e.g. the in-vivo prediction of flap tissue viability [11]. Often low spatial resolution images are adequate, only necessitating registration techniques of relatively low sophistication to compensate for gross movements of the subject during data acquisition, However, this does not suffice for applications where a high degree of spatial resolution is called for. For example, in dermatological investigations, the inevitable presence of motion artefacts is cited as a major limitation for NIR imaging [12]. As such, we anticipate that significant improvements in the recovery of localised skin chromophore features are attainable from accurate image registration.

NIR images (at 10 nm increments over the wavelength range 650 – 1050 nm) were obtained from the NIR hyperspectral datacube (256 x256x41) of a healthy, 54 year old, male volunteer's face [13]. Image #20 in the datacube was selected as a reference image. Two Gabor filters, tuned to (U=0.1, V=0.00, σ=2.0) and (U=0.00, V=0.1, σ=2.0), were used to compute LF representation of all images. The size of the real filter was 5x5, imaginary 9x9.

The NIR parameter images of oxygen saturation and deoxy-hemoglobin, derived following registration with the PA are shown in Figs. 3 and 4, respectively. For comparison and validation each figure also presents the unregistered image and its registered counterpart using an established registration technique (maximization of Mutual Information) [14]. Considering the oxygen saturation image one can clearly see that registration has recovered a localized area (circled), which was previously obscured. From visual image records we see that this feature corresponds to an epidermal pigmented lesion, tentatively assigned as a "Cafe-au-Lait Macule". The same feature is also well resolved in the registered oxy-hemoglobin and total hemoglobin images (not shown). Prior to registration the deoxy-hemoglobin image exhibits an unremarkable distribution with no localized features distinguished. However, an entirely different feature from above is revealed in the registered image (circled). This faint, highly localized feature corresponds visually to the location of a small closed comedo, a type of non-inflammatory acne lesion. Also of note are that images derived from the PA and reference techniques give very similar results, although both introduce a slight blurring to certain features. This is especially noticeable for the eyes and is attributed to non-linear distortions, prevalent in changes of facial expression occurring during the relatively long data acquisition, that tend to confound the current algorithm that uses an affine registration model. It is anticipated that this limitation may be overcome, with a registration program that corrects for general distortions.

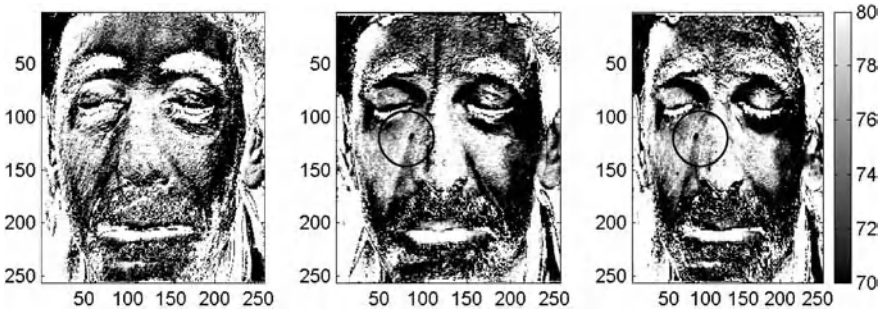


Fig. 3. Image of the oxygen saturation distribution derived from (left) the unregistered NIR hyperspectral datacube; (centre) the same image after registration has been performed using an algorithm based on a maximization of Mutual Information; (right) the same image after registration using the Patch algorithm. The feature recovered (circled) corresponds visually to an epidermal pigmented lesion (Cafe-au-Lait Macule).

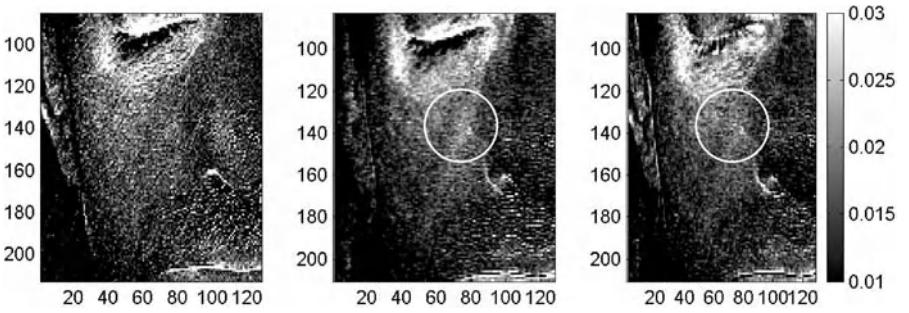


Fig. 4. Detail of the deoxy-hemoglobin distribution derived from (left) the unregistered NIR hyperspectral datacube; (centre) the same image after registration has been performed using an algorithm based on a maximization of Mutual Information; (right) the same image after registration using the Patch algorithm. The feature recovered (circled) corresponds visually to the location of a closed comedo acne lesion.

Since the subject was a healthy control these results confer little diagnostic significance, however, it does provide an unambiguous demonstration that PA registration of the NIR hyperspectral datacube can recover previously unresolved, localized dermal features.

In conclusion, mapping images of different modality (T_1/T_2, NIR hyperspectral datacube) to LF one can register them using the PA, a unimodal registration program. Constructing a suitable LF representation of the images, i.e., fine tuning of Gabor filters is crucial in such approach. In several attempts, when filters were not tuned properly, the registration simply failed to register the LF images. We found that the LF images are sensitive to the size of Gabor filter, e.g., for the T_1/T_2 images mentioned above the increase of the size of the real part of the filter from 5 to 7 may result in completely different LF images which often is not suitable for registration at all.

Acknowledgements

The authors would like to thank Dr. B.C. Vemuri of University of Florida for his suggestions on LF representation of images.

References

1. Viola, P., Wells W.M.: Alignment by maximization of mutual information. International Journal of Computer Vision **24** (1997) 137-154
2. Fleet, D.J., Jepson, A.D.: Computation of component image velocity from local phase information. International Journal of Computer Vision **5** (1990) 77-104
3. Liu, J., Vemuri, B.C., Marroquin, J.L.: Local frequency representations for robust multimodal image registration. IEEE Trans. Medical Imaging, 21 (2002) 462-469
4. Alexander, M.E.: Fast hierarchical noniterative registration algorithm, International Journal of Imaging Systems and Technology, Vol. 10 (1999), pp. 242-257
5. Zhilkin P., Alexander M.E.: 3D image registration using a fast noniterative algorithm. Magnetic Resonance Imaging 18 (2000) 1143-1150
6. Zhilkin P., Alexander M.E.: Nonlinear Registration of 3D Images Using Patch Algorithm. ISMRM, May 18-24, 2002, Honolulu, Hawaii, USA
7. Liu J., Vemuri B.C., Bova F.: Efficient multi-modal image registration using local-frequency maps. Machine Vision and Applications 13 (2002) 149-163
8. Thévenaz P., Unser M.: Optimization of Mutual Information for Multiresolution Image Registration. IEEE Trans. Image Processing, 9 (2000) 2083-2099
9. http://www.bic.mni.mcgill.ca/brainweb/
10. Cocosco C.A., Kollokian V., Kwan R.K.-S., Evans A.C.: BrainWeb: Online Interface to a 3D MRI Simulated Brain Database, NeuroImage, 5(4) 1977, part 2/4, S425 -- Proceedings of 3-rd International Conference on Functional Mapping of the Human Brain, Copenhagen, May 1997
11. Stranc M.F., Sowa M.G., Abdulrf B. and Mantsch H.H.: Assessment of tissue viability using near-infrared spectroscopy, Br. J. Plast. Surg, 51(3) (1998) 210 – 217
12. Kollias N. and Stamatas G.N.: Optical non-invasive approaches to diagnosis of skin diseases, Journal of investigative dermatology, The Symposium proceedings the Society for Investigative Dermatology, Inc. and European Society for Dermatological Research, 7(1) (2002) 64 – 75
13. Mansfield C.D, and Alexander M.E.: Registration of Thermal and Near Infrared Hyperspectral Datacubes for Recovery of Localized Thermal or Hemodynamic Features, manuscript submitted to IEEE Transactions on Medical Imaging (2003)
14. Alexander M.E. and Summers A.R.: Fast Registration Algorithm Using a Variational Principle for Mutual Information, Proceedings of SPIE Medical Imaging Conference, San Diego, Feb 15-20, 2003, to appear

FLIRT: A Flexible Image Registration Toolbox

Bernd Fischer and Jan Modersitzki

Institute of Mathematics
University of Lübeck, 23560 Lübeck, Germany
{fischer,modersitzki}@math.uni-luebeck.de
http://www.math.uni-luebeck.de

Abstract. Image registration is central to many challenges in medical imaging today and has a vast range of applications. The purpose of this note is to provide a toolbox for intensity based non-rigid registration problems. To do so, we review some of the most promising non-linear registration strategies currently used in medical imaging and show that all these techniques may be phrased in terms of a variational problem and allow for a unified treatment.
Depending on the application at hand, it is often desirable to constrain the wanted deformation. The idea is to incorporate higher level information about the expected deformation. We examine the most common constraints and show again that they may be conveniently phrased in a variational setting.
As a consequence, all of discussed modules allow for fast implementations and may be combined in any favorable order. We discuss individual methods for various applications, including the registration of histological serial sections of a human brain.

1 Introduction

In the last two decades, computerized non-rigid image registration has played an increasingly important role in medical imaging, see for example Maintz & Viergever [14], Fitzpatrick et al. [11], and references therein. The problem of registration arises whenever images acquired from different subjects, at different times, or from different scanners need to be combined for analysis or visualization.

Due to the wide range of applications a variety of different registration techniques has been developed. Here, we focus on so-called *intensity-driven approaches*. These schemes aim to match intensity patterns between a deformed scan and the target based on a rigorous mathematical criterion. Depending on the application, different strategies may be employed. From a practical point of view, it is desirable to incorporate properties of the underlying problem into the registration scheme. Here, we provide a toolbox of registration routines which enables the user to choose in a consistent way building blocks for schemes which cover a wide range of applications. The idea is to phrase each individual block in terms of a variational formulation. This not only allows for a unified treatment but also for fast and reliable implementation. The various building blocks

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 261–270, 2003.

comprises three categories: smoother and internal forces, distances and external forces, and "hard" or "soft" constraints. The *internal forces*, are defined for the wanted *displacement field* itself and are designed to keep the displacement field smooth during deformation. In contrast, the *external forces* are computed from the image data and are defined to drive the displacement in order to arrive at the desired registration result. Whereas the internal forces implicitly constrain the displacement to obey a smoothness criterion, the additional *constraints* force the displacement to satisfy explicit criteria, like for example landmark or volume preserving imposed constraints.

The paper is organized as follows. In Section 2 we summarize the most popular choices for the above outlined building blocks. Furthermore, we set up a general and unified framework for automatic non-rigid registration. In Section 3 we show in more detail, how these building blocks can be translated into a variational setting. It is this formulation, which allows for a fast and reliable numerical treatment. In Section 4 we indicate on how to actually implement the registration schemes. An example in Section 5 highlights the importance of having more than one regularizer at hand.

2 A Flexible Image Registration Toolbox

Registration is the determination of a geometrical transformation that aligns points in one view of an object with corresponding points in another view of the same object or a similar object. There exist many instances in a medical environment which demand for a registration, including the treatment verification of pre- and post-intervention images, study of temporal series of cardiac images, and the monitoring of the time evolution of an agent injection subject to patient motion. Another important area is the need for combining information from multiple images, acquired using different modalities, like for example computer tomography (CT) and magnetic resonance imaging (MRI).

To be successful, each individual application should be treated by a specific registration technique. It is the purpose of this note to provide a toolbox for nonlinear registration schemes, which may be adapted to the special problem class under consideration. The main building blocks of this toolbox resemble typical user demands and may be assembled in a consistent and intuitive fashion.

Given two images, a reference R and a template T, the aim of image registration is to find a global and/or local transformation from T onto R in such a way that the transformed template matches the reference. Ideally there exists a coordinate transformation u such that the reference R equals the transformed template T_u. Given such a displacement u, the registration problem reduces to a simple interpolation task. However, in general it is not possible to come up with a perfect u, and the registration problem is to compute an application conformal transformation u, given the reference and template image.

It should be pointed out, that apart from the fact that a solution may not exist, it is not necessarily unique. For an example, see Modersitzki [15]. In other words, intensity based registration is inherently an ill-posed problem.

A displacement u which does produce a perfect or nearly perfect alignment of the given images is not necessarily a "good" displacement. For example, a computed displacement which interchanges the eyes of one patient when registered to a probabilistic atlas in order to produce a nearly perfect alignment, has obviously to be discarded. Also, folding and cracks introduced by the displacement are typically not wanted. Therefore it is desirable to have a possibility to incorporate features into the registration model, such that the computed displacement u does resemble the properties of the acquisition, like for example the elastic behavior of a human brain. To mimic the elastic properties of the objects under consideration is a striking example for internal forces. These forces constrain the displacement to physically meaningful movements.

In contrast, the external forces are designed to push the deformable template into the direction of the reference. These forces are based upon the intensities of the images. The idea is to design a similarity measure, which is ideally calculated from all voxel values. An intuitive measure is the sum of squares of intensity differences (SSD). This is a reasonable measure for some applications like the serial registration of histological sections. If the intensities of corresponding voxels are no longer identical, the SSD measure may perform poorly. However, if the intensities are still linearly related, a correlation (CC) based measure is the measure of choice for monomodal situations. In contrast, the mutual information (MI) related measure is based on the cooccurrence of intensities in both images as reflected by their joint intensity histogram. It appears to be the most successful similarity measure for multimodal imaginary, like MR-PET; cf. e.g., Roche [16] or Viola [19].

Finally, one may want to guide the registration process by incorporating additional information which may be known beforehand. Among these are landmarks and fiducial markers. Sometimes it is also desirable to impose a local volume-preserving (incompressibility) constraint which may, for example, compensate for registration artifacts frequently observed by processing pre- and post-contrast images. Depending on the application and the reliability of the specific information, one may want to insist on a perfect fulfilment of these constraints or on a relaxed treatment. For examples, in practise, it is a tricky (and time consuming) problem to determine landmarks to subvoxel precision. Here, it does not make sense to compute a displacement which produces a perfect one to one match between the landmarks.

Summarizing, the general registration problem may be phrased as follows.

(IR) image registration problem:

$$\mathcal{J}[u] = \mathcal{D}[R, T; u] + \alpha \mathcal{S}[u] = \min,$$
$$\text{subject to} \quad \mathcal{C}_j[u] = 0, \quad j = 1, 2, \ldots, m.$$

Here, $\mathcal{D}$ models the distance measure (external force, e.g., MI), $\mathcal{S}$ the smoother (internal force, e.g., elasticity), and $\mathcal{C}$ explicit constraints (e.g., landmarks). The regularization parameter α may be used to control the strength of the smoothness

of the displacement versus the similarity of the images. In the following we will discuss these building blocks in more detail.

3 Toolbox Building Blocks

Our approach is valid for images of any spatial dimension d, i.e., there is no restriction to $d = 2, 3, 4$. The reference and template images are represented by the compactly supported mappings $R, T : \Omega \to \mathbb{R}$, where without loss of generality, $\Omega =]0, 1[^d$. Hence, $T(x)$ denotes the intensity of the template at the spatial position x, where for ease of discussion we set $R(x) = b_R$ and $T(x) = b_T$ for all $x \notin \Omega$. Here, b_R and b_T are appropriately chosen background intensities. The overall goal is to find a *displacement* u, such that ideally T_u is similar to R, where T_u is the deformed image, i.e., $T_u(x) = T(x - u(x))$. Note that $u = (u_1, \ldots, u_d)$ denotes a vector field.

The starting point of our numerical treatment is the minimization of problem (IR). In order to compute a minimizer we apply a steepest descent method, where we take advantage of the calculus of variations. To end up with an efficient and fast converging scheme, we require to have explicit expressions of the derivatives of building blocks $\mathcal{D}$, $\mathcal{S}$, and $\mathcal{C}$. In the following subsections we will exemplarily discuss the most popular building blocks as well as their derivatives.

Smoother and Internal Forces. The nature of the deformation depends strongly on the application under consideration. For example, a slice of a paraffin embedded histological tissue does deform elastically, whereas the deformation between the brains of two different individuals is most likely not elastically. Therefore, it is necessary to supply a model for the nature of the expected deformation.

We now present some of the most prominent smoothers $\mathcal{S}$ and discuss exemplarily the GÂTEAUX-derivatives for two of them. An important point is, that we are not restricted to a particular smoother $\mathcal{S}$. Any smoother can be incorporated into this toolbox, as long as it possesses a GÂTEAUX-derivative.

In an abstract setting, the GÂTEAUX-derivative looks like

$$d\mathcal{S}[u; v] := \lim_{h \to 0} \frac{1}{h}(\mathcal{S}[u + hv] - \mathcal{S}[u]) = \int_\Omega \langle \mathcal{A}[u], v \rangle_{\mathbb{R}^d} \; dx,$$

where $\mathcal{A}$ denotes the associated linear partial differential operator. Note that for a complete derivation one also has to consider appropriate boundary conditions. However, these details are omitted here for presentation purposes; see MODERSITZKI [15] for details.

Elastic Registration. This particular smoother measures the elastic potential of the deformation. In connection with image registration it has been introduced by BROIT [3] and discussed by various image registration groups; see, e.g., BAJCSY & KOVAČIČ [2] or FISCHER & MODERSITZKI [7]. The partial differential

operator is the well-known NAVIER-LAMÉ operator. For this smoother, two natural parameters, the so-called LAMÉ-constants can be used in order to capture features of the underlying elastic body. A striking example, where the underlying physics suggests to look for deformations satisfying elasticity constraints, is the three-dimensional reconstruction of the human brain from a histological sectioning. Details are given in SCHMITT [18] and MODERSITZKI [15].

Fluid Registration. Due to the fact that an elastic body memorizes its non-deformed initial state (rubber band), elastic registration schemes are only able to compensate for small deformations. The situation changes for the viscous fluid model. Here the body adapts to its current state (honey) and consequently is much more flexible than an elastic body. The viscous fluid approach was introduced to image registration by CHRISTENSEN [4]. His derivation was based on a specific linearization of the NAVIER-STOKES equation. However, there is yet another derivation of the underlying partial differential equations, which does fit into "design rules" of our toolbox. Roughly speaking, one obtains these equations by considering the elastic potential of the *velocity* of the displacement field. It should come as no surprise that the partial differential operator is again the NAVIER-LAMÉ operator, this time, however, applied to the velocity. The wanted deformation is related to the velocity via the material derivative and is straightforward to recover.

Since the viscous fluid approach is quite flexible, it is mainly used when the focus is more on similarity than on a "natural deformation process". For example, for the design of a probabilistic brain atlas, a biophysical model for the nature of the deformations is not available, but the fluid registration has been proven to be a valuable tool; cf., e.g. D'AGOSTINO ET AL. [5].

Diffusion Registration. For image registration problems FISCHER & MODERSITZKI [8] introduced the so-called diffusion regularization

$$\mathcal{S}^{\text{diff}}[u] := \frac{1}{2} \sum_{\ell=1}^{d} \int_{\Omega} \|\nabla u_\ell\|^2 \, dx, \tag{1}$$

which is well-known for *optical flow* applications; see HORN & SCHUNCK [13]. The associated GÂTEAUX-derivative leads to the well-studied LAPLACE- operator, i.e., $\mathcal{A}^{\text{diff}}[u] = \Delta u = (\Delta u_1, \ldots, \Delta u_d)$, where $\Delta u_\ell = \partial_{x_1 x_1} u_\ell + \cdots + \partial_{x_d x_d} u_\ell$. The main reason for introducing this smoother was its exceptional computational complexity. FISCHER & MODERSITZKI [8] devised an $\mathcal{O}(N)$ (!) implementation of the registration scheme, where N denotes the number of image voxels. It is based on an additive operator splitting scheme (which parallelizes in a very natural way). Its outstanding computational speed makes the diffusion registration scheme to a very attractive option for high-resolution, high dimensional, and/or time critical applications. Examples include the registration of a time series of three-dimensional MRI's or the online correction of the so-called brain shift during surgery.

Curvature Registration. As a last example, we present the curvature smoother

$$\mathcal{S}^{\text{curv}}[u] := \tfrac{1}{2}\sum_{\ell=1}^{d}\int_\Omega (\Delta u_\ell)^2 \; dx, \tag{2}$$

introduced by Fischer & Modersitzki [9]. The design principle behind this choice was the idea to make the non-linear registration phase more robust against a poor (affine linear) pre-registration. Since the smoother is based on second order derivatives, affine linear maps do not contribute to its costs, i.e.,

$$\mathcal{S}^{\text{curv}}[Cx+b] = 0, \quad \text{for all} \quad C \in \mathbb{R}^{d\times d},\ b \in \mathbb{R}^d.$$

In contrast to other non-linear registration techniques, affine linear deformations are corrected naturally by the curvature approach. This advantage is illustrated by Figure 1, where the results of a fluid and a curvature based registration of two X-ray images are compared; see also Section 5. Again the Gâteaux-derivative is explicitly known and leads to the so-called bi-harmonic operator $\mathcal{A}^{\text{curv}}[u] = \Delta^2 u$.

Distances and External Forces. Another important building block is the similarity criterion. As for the smoothing operators, we concentrate on those measures $\mathcal{D}$ which allow for differentiation. Moreover, we assume that there exists a function $f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$, the so-called *force field*, such that

$$\begin{aligned} d\mathcal{D}[R,T;u;v] &= \lim_{h\to 0}\frac{1}{h}(\mathcal{D}[R,T;u+hv] - \mathcal{D}[R,T;u]) \\ &= \int_\Omega \langle f(R,T,x,u(x)), v(x)\rangle_{\mathbb{R}^d} \; dx. \end{aligned}$$

Again, we are not restricted to a particular distance measure. Any measure can be incorporated into our toolbox, as long as it permits a Gâteaux-derivative.

The most common choices for distance measures in image registration are the *sum of squared differences*, *cross correlation*, *cross validation*, and *mutual information*. We give explicit formulae for only two of them; for more information see, e.g., Modersitzki [15] or Roche [16].

Sum of Squared Differences. The measure is based on a point-wise comparison of image intensities,

$$\mathcal{D}^{\text{SSD}}[R,T;u] := \tfrac{1}{2}\int_\Omega \big(R(x) - T_u(x)\big)^2 \; dx,$$

and the force-field is given by $f^{\text{SSD}}(R,T,x,y) = (T(x-y) - R(x))\cdot\nabla T(x-y)$. This measure is often used when images of the same modality have to be registered.

Mutual Information. Another popular choice is mutual information. It basically measures the entropy of the joint density $\rho^{R,T}$, where $\rho^{R,T}(g_1,g_2)$ counts the number of voxels with intensity g_1 in R and g_2 in T. The precise formula is

$$\mathcal{D}^{\text{MI}}[R,T;u] := -\int_{\mathbb{R}^2} p^{R,T_u} \log \tfrac{p^{R,T_u}}{p^R p^{T_u}} \; d(g_1,g_2),$$

where p^R and p^{T_u} denote the marginal densities. Typically, the density is replaced by a PARZEN-window estimator; see, e.g. VIOLA [19]. The associated force-field is given by

$$f^{\mathrm{MI}}(R,T,x,y) = \int_\Omega [\Psi_\sigma * \partial_{g_2} L^{R,T_u}](R(x),T_u(x)) \cdot \langle \nabla T_u(x), v(x)\rangle_{\mathbb{R}^d} ,$$

where $L^{R,T_u} := 1 + p^{R,T_u}(\log p^{R,T_u} - \log(p^R p^{T_u})$ and Ψ is the PARZEN-window function; see, e.g., HERMOSILLO [12] or D'AGOSTINO ET AL. [6]. This measure is useful when images of a different modality have to be registered.

Additional Constraints. Often it is desirable to guide the registration process by incorporating additional information which may be known beforehand, like for example markers. To incorporate such information, the idea is to add additional constraints to the minimization problem. For example, to restrict the deformation to volume preserving mappings, one has to add the quantity

$$\mathcal{C}[u] := \frac{1}{2}\int_\Omega (\det \nabla u)^2 \; dx$$

to the smoother; see also ROHLFING & MAURER [17]. Note that the JACOBIAN $\det \nabla u(x)$ has to vanish, if the deformation at x is incompressible.

In other applications, one may want to incorporate landmarks or fiducial markers. Let r^j be a landmark in the reference image and t^j be the corresponding landmark in the template image. The toolbox allows for either adding explicit constraints

$$\mathcal{C}_j[u] := u(t^j) - t^j + r^j, \quad j = 1,2,\ldots,m,$$

which have to be precisely fulfilled $\mathcal{C}_j[u] = 0$ ("hard" constraints), or by adding an additional cost term

$$\mathcal{C}[u] := \sum_{j=1}^m \lambda_j \, \|\mathcal{C}_j[u]\|^2_{\mathbb{R}^d}$$

to the smoother ("soft" constraints, since we allow for deviations). For a more detailed discussion, we refer to FISCHER & MODERSITZKI [10].

4 Numerical Treatment

As already pointed out, our numerical approach is based on the EULER- LAGRANGE equations for the problem (IR)

$$\mathcal{A}[u](x) + f(R,T,x,u(x)) + \sum_{j=1}^m \lambda_j d\mathcal{C}_j[u](x) = 0 \quad \text{and} \quad \mathcal{C}_j[u] = 0, \; j = 1,\ldots,m,$$

where the λ_j's are LAGRANGE parameter. Roughly speaking, all associated GÂTEAUX-derivatives have to vanish. It remains to efficiently solve this system of non-linear partial differential equations. After invoking a time-stepping

approach and after an appropriate space discretization, we finally end up with a system of linear equations. As it turns out, these linear systems have a very rich structure, which allows one to come up with very fast and robust solution schemes for all of the above mentioned building blocks. It is important to note that the system matrix does not depend on the force field and the constraints. Thus, changing the similarity measure or adding additional constraints does not change the favorable computational complexity. Moreover, fast and parallel solution schemes can be applied to even more reduce the computation time.

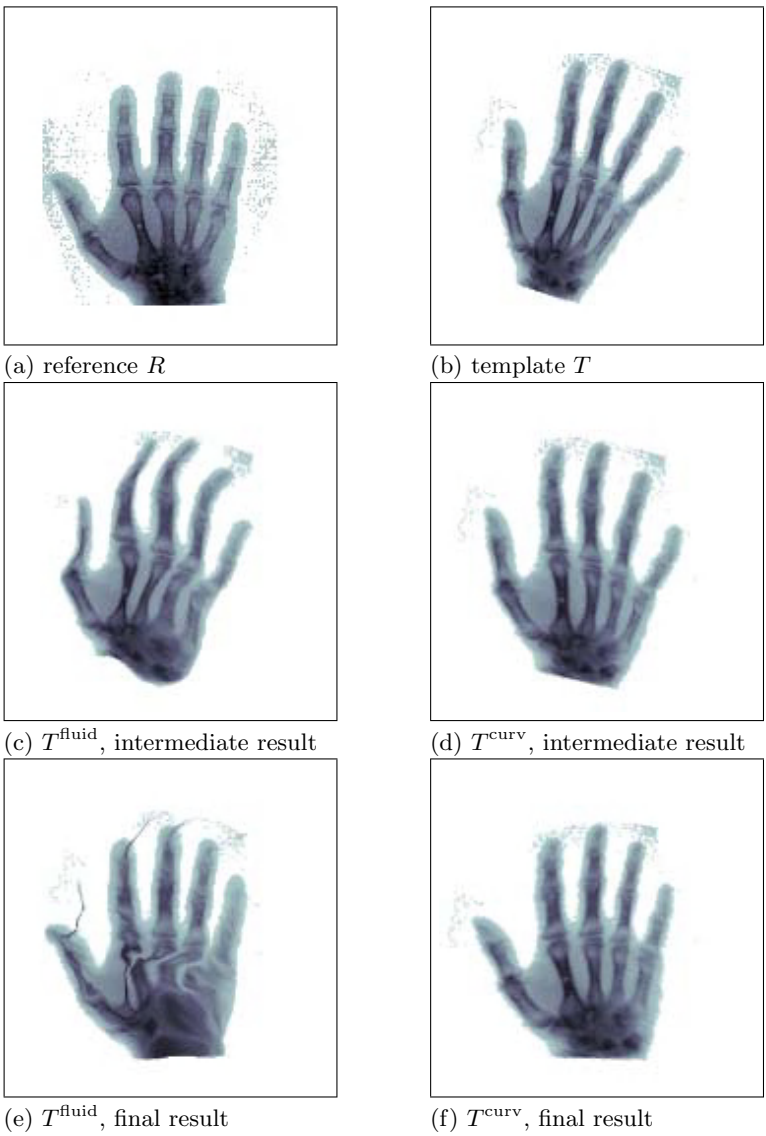
5 An Example: X-Rays of Hands

We present a synthetic example in order to demonstrate the fact, that changing the smoother may dramatically affect the registration result. Here, we compare the fluid and curvature smoother, both accompanied with the SSD measure. In Figure 1, a reference (a) and a template image (b) are displayed (modified X-rays from human hands, images from AMIT [1]). Obviously, an affine linear pre-registration (rotation of about 45 degrees and re-scaling) would improve the similarity of the images considerably. However, in order to keep the issue of interest clear, we did not apply any pre-registration. For the fluid registration we end up with the deformed template displayed in Figure 1(e) and for the curvature registration we obtain the result shown in Figure 1(f) (Figure 1(c,d) show intermediate results of the time-stepping scheme). As it is apparent from this example, the fluid approach produced a miss-registration whereas the curvature approach produced a satisfactory result.

The main point is that the fluid registration (as well as the other approaches) does penalize affine linear deformations and may therefore privilege non-linear deformations, as is clearly visible in Figure 1(c). Due to its flexibility, the fluid method finally recovers the reference, where, however, the deformation field is unnatural. In contrast, the curvature approach does not penalize affine linear deformations as can be seen in Figure 1(d), which displays almost a rotated and re-scaled template image. Note that the deformation is not completely linear and note that this is an extreme example. For comparison reasons, we applied non-optimized methods on a single scale (though all building blocks can be applied in a multiscale resolution as well).

6 Conclusions

In this note we presented a general approach to image registration. Its flexibility enables one to integrate and to combine in a consistent way various different registration modules.We discussed the use of different smoothers, distance measures, and additional constraints. The numerical treatment is based on the solution of a partial differential equation related to the EULER-LAGRANGE equations. These equations are well studied and allow for fast, stable, and efficient schemes. Due to page limits, we reported on only one example, showing the effect of different



(a) reference R (b) template T

(c) T^{fluid}, intermediate result (d) T^{curv}, intermediate result

(e) T^{fluid}, final result (f) T^{curv}, final result

Fig. 1. Two modified X-ray images of human hands; see also Amit [1]. (a) reference image, (b) template image, (c) intermediate result of fluid registration, (d) intermediate result of curvature registration, (e) final result of fluid registration, and (f) final result of curvature registration.

smoothers. We will report in a forthcoming paper on an exhaustive comparison of the various building blocks.

Part of the software is available via `http://www.math.uni-luebeck.de/SAFIR`.

References

1. Y. Amit, *A nonlinear variational problem for image matching*, SIAM J. Sci. Comput. **15**(1), 207–224, 1994.
2. R. Bajcsy and S. Kovačič, *Multiresolution elastic matching*, Computer Vision, Graphics and Image Processing **46**, 1–21, 1989.
3. C. Broit, *Optimal registration of deformed images*, Ph.D. thesis, Computer and Information Science, Uni Pensylvania, 1981.
4. G. E. Christensen, *Deformable shape models for anatomy*, Ph.D. thesis, Sever Institute of Technology, Washington University, 1994.
5. E. D'Agostino, F. Maes, D. Vandermeulen, P. Suetens. A viscous fluid model for multimodal non-rigid image registration using mutual information. *MICCAI 2002*, Tokyo, 23-26 September, 2002.
6. E. D'Agostino, J. Modersitzki, F. Maes, D. Vandermeulen, B. Fischer, and P. Suetens, *Free-form registration using mutual information and curvature regularization*, Preprint A-03-05, Institute of Mathematics, University of Lübeck, 2003.
7. B. Fischer and J. Modersitzki, *Fast inversion of matrices arising in image processing*, Num. Algo. **22**, 1–11, 1999.
8. B. Fischer and J. Modersitzki, *Fast diffusion registration*, AMS Contemporary Mathematics, Inverse Problems, Image Analysis, and Medical Imaging **313**, 117–129, 2002.
9. B. Fischer and J. Modersitzki, *Curvature based image registration*, JMIV **18**(1), 81–85, 2003.
10. B. Fischer and J. Modersitzki, *Combination of automatic non-rigid and landmark based registration: the best of both worlds*, Preprint A-03-01, Institute of Mathematics, University of Lübeck, 2003.
11. J. M. Fitzpatrick, D. L. G. Hill, and C. R. Maurer Jr., *Image registration*, Handbook of Medical Imaging, Volume 2: Medical Image Processing and Analysis (M. Sonka and J. M. Fitzpatrick, eds.), SPIE, 447–513, 2000.
12. G. Hermosillo, *Variational methods for multimodal image matching*, Ph.D. thesis, Université de Nice, France, 2002.
13. B. K. P. Horn and B. G. Schunck, *Determining optical flow*, Artificial Intelligence **17**, 185–204, 1981.
14. J. B. Antoine Maintz and M. A. Viergever, *A survey of medical image registration*, Medical Image Analysis **2**(1), 1–36, 1998.
15. J. Modersitzki, *Numerical methods for image registration*, Oxford University Press, to appear 2003.
16. A. Roche, *Recalage d'images médicales par inférence statistique*, Ph.D. thesis, Université de Nice, Sophia-Antipolis, France, 2001.
17. T. Rohlfing and C. R. Maurer, Jr, *Volume-Preserving Non-Rigid Registration of MR Breast Images Using Free-Form Deformation with an Incompressibility Constraint*, IEEE TMI, to appear 2003.
18. O. Schmitt, *Die multimodale Architektonik des menschlichen Gehirns*, Habilitation (in German), Insitute of Anatomy, Medical University of Lübeck, Germany, 2001.
19. P. A. Viola, *Alignment by maximization of mutual information*, Ph.D. thesis, Massachusetts Institute of Technology, 1–155, 1995.

Narrow Band to Image Registration in the Insight Toolkit

Lydia Ng[1] and Luis Ibáñez[2]

[1] Insightful Corporation, Seattle, WA, USA
lng@insightful.com
http://www.insightful.com
[2] Kitware Inc., Clifton Park, NY, USA
luis.ibanez@kitware.com
http://www.kitware.com

Abstract. This paper introduces the new concept of narrow-band to image registration. Narrow-banding is a common technique used in the solution of level set approaches to image processing. For our application, the narrow-band describes the shape of an object by using a data structure containing the signed distance values at a small band of neighboring pixels. This compact representation of an object is well suited for performing registration against a standard image as well as against another narrow-band. The novel technique was implemented in the registration framework of the NLM Insight Toolkit (ITK). This implementation illustrates the great advantage of a modular framework structure that allows researchers to concentrate in the interesting aspects of a new algorithm by building on an existing set of predefined components for providing the rest of standard functionalities that are required.

1 Overview

1.1 Level Sets and Narrow Bands

Level sets have proved in recent years to be a powerful approach for solving image processing problems. Among the many applications of level sets to image processing, the task of segmentation has attracted significant attention. Typically, the segmentation process involves the evolution of an initial contour. In the level set framework, the contour is not manipulated directly. Rather, the contour is embedded in a higher order function known as the level set function, $\phi(\mathbf{x})$. The evolving contour is defined by the zero level set, $\phi(\mathbf{x}) = 0$, with $\phi(\mathbf{x}) < 0$, representing the interior and $\phi(\mathbf{x}) > 0$, representing the exterior. The evolution of the level set is typically governed by a partial differential equation (PDE).

Some of the major advantages of using level sets are that arbitarily complex shapes (e.g with corners and protrusions) can be modelled and topological changes such as merging and splitting of regions are handled implicitly. Further, by selecting different configurations of the PDE, it is possible to endow the contour with particular properties like smoothness, attraction to image edges [1,2] and the ability to partition an image based on region statistics [3], just to name a few. A thorough review of level sets methods can be found in [4]. One of the drawbacks of level set techniques is the intense

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 271–280, 2003.

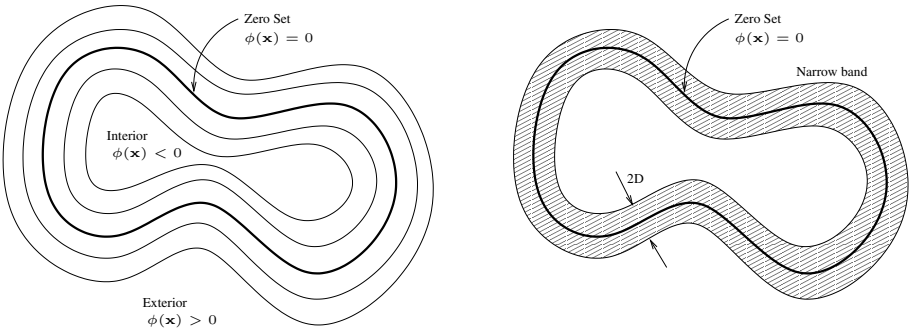


Fig. 1. Representation of the level set function, zero set and narrow-band concepts in $2D$.

computational load due to the iterative updates of the entire level set function. For the purpose of segmentation, usually the only region of interest is the small neighborhood about the zero level set. This observation lead to the development of narrow-banding techniques. The idea is that, all the interesting activities during the evolution of the level set happen at a certain distance D of the zero set, forming a band of width $2D$. By solving the PDE only in the domain of the narrow-band, instead of the entire level set function, a great reduction in computational load is achieved.

Figure 1 illustrates the basic concept of level sets, the zero set and the narrow-band. The curves at the left represent iso-contours of $\phi(\mathbf{x})$. The zero set is highlighted as a bold curve . In this particular case, the zero set represents the shape of a contour in 2D. On the right, a narrow-band is represented as the region composed of all the points at less than D distance from the zero set.

1.2 Motivation for Using A Narrow-Band for Registration

Image registration is an active research area. Methods using only image intensities (e.g. mean squares, normalized correlation and mutual information) have been well documented [5,6]. Methods using distance information from edge features have also been demonstrated [7]. The motivation for using narrow-band shape representation for performing registration is to combine the benefits of both schemes. Narrow band to image registration can be used for many applications. One example is to find a particular template shape in an image. Another example is to use it in a two step image to image registration. A structure of interest is segmented from one image, then a narrow-band containing a signed distance to the structure boundary is created. Registration is then performed using narrow-band information and intensity of the other image.

There are several advantages to using narrow-band to image registration:

– The use of a narrow-band shape representation of a structure, focuses the registration process to use meaningful information as the basis of the registration. This is not always the case in intensity based registration. It is often difficult to establish what are the sources of the main contributions to the goodness of fit measure

to be optimized during the registration process. For example, a large expanse of background, variation in the field of view or extraneous objects with high gradient features (e.g. surgical instruments in image guided surgery applications) can have undesirable effects.
- The narrow-band is a compact representation of the structure because only pixels close to the object boundary are used. This leads to a reduction in the memory requirements as well as the computational load of the registration process. For metrics like mean squares and normalized correlation, this makes a difference of one to two order of magnitude depending on the tortuosity of the anatomical structure boundary.
- The narrow-band not only represents the shape contour but also a transition zone around it. This results in a large capture radius in the optimization space. The capture radius can in fact be regulated by selecting the different widths of the narrow-band.
- The configuration of a narrow-band strongly resembles the sensitivity profile of biological sensors, such as the cones in the retina. Those sensors have a central excitatory region surrounded by an inhibitory region [8]. This configuration provides a balance between sensitivity and precision of signal location [9]. This resemblance provides an intuitive motivation for further exploring this technique. The results presented in section 4 provide encouraging support in favor of this intuition.

2 Implementation in ITK

We implemented the narrow-band to image registration scheme using the NLM Insight Toolkit or ITK (www.itk.org). ITK is an open-source cross-platform C++ toolkit containing many components and frameworks for performing registration and segmentation for medical images [10]. In particular, ITK contains various methods for performing the initial segmentation of the structure, classes to represent and generate narrow-bands and a flexible and extensible framework to create registration processes. Additionally, the templated environment allow processing of images of different pixel types and most algorithms are implemented to support two, three and higher dimensional images.

2.1 Narrow Band Representation

In ITK, a narrow-band data structure consists of an array of node objects representing the pixels of the narrow-band. Each node contains the position of the pixel it represents and the level set value at that pixel. Figure 2 illustrates how ITK image filters can be used to create a narrow-band structure for our narrow-band to image registration application. First, the original image is passed to a segmentation filter to produce a binary image corresponding to the spatial support of the anatomical structure of interest. A review of segmentation methods is beyond the scope of this paper. Many segmentation methods are available in ITK. We will suppose that an adequate algorithm is used. In this example, the left venticle is extracted using the geodesic active contours [2] filter implemented in ITK. The binary image is then passed to a filter that generates a signed distance map from the structure using the fast marching reinitialization method

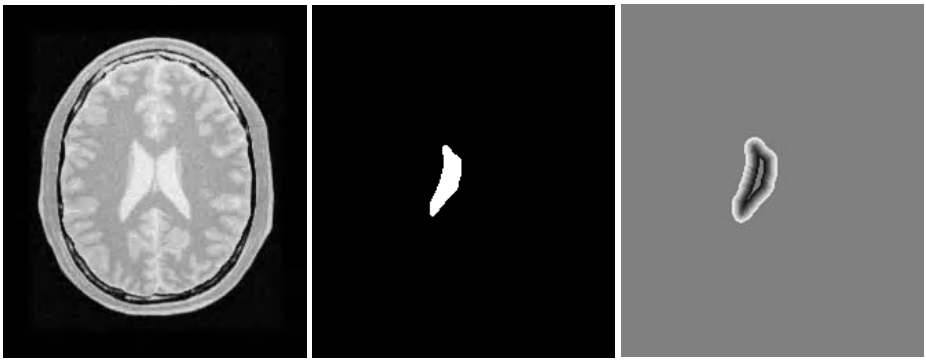


Fig. 2. Left, input image. Center, segmentation of the left ventricle from the fixed image using the geodesic active contours filter. Right, narrow-band representing the ventricle.

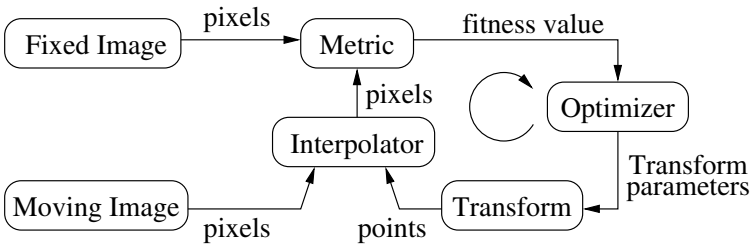


Fig. 3. Modular components involved in a image to image registration process in ITK.

described in [4]. The narrow-band structure is initialized using the distance map to contain nodes which are within a user defined band width from the zero set. Note that, the shape does not necessarily have to come from an image, all that is required is a binary representation of the shape to create a narrow-band. The image on the right is a representation of the narrow-band. In this 8-bit image, the zero of the level set is represented with the medium gray level of 128. Bright pixels represent positive values in the narrow-band. Dark pixels represent negative values in the narrow-band. The gray background in this image is not part of the narrow-band structure.

2.2 Registration Framework of the Insight Toolkit

The Insight toolkit implements a modular approach for solving registration problems. All the components involved in the process of registration have been separated by their functionality. For each functional component different implementations are offered in the toolkit. A programming interface is also defined in such a way that a user can add his/her own components and still take advantage of all the others predefined in the toolkit. Figure 3 shows the major components of a registration problem as defined in ITK. The two images to be registered are called *fixed* and *moving*. A spatial transform will be used to map points from the space of the fixed image to the space of the moving

image. Since most of this mappings will fall on non-grid positions of the moving image, an interpolator is required for computing intensity values at those positions. A metric is defined in order to evaluate the degree at which the fixed image matches the intensity levels of the transformed moving image. This metric acts as the cost function to be optimized and is passed to a generic optimizer. Since the goal of registration is to find the transform that optimizes the metric, the search space for the optimizer is necessarily defined by the set of parameters in the transform.

Typically the two input components of registration are images. However, nothing prevents developers from extending the concepts in the basic diagram of Figure 3. The novel technique presented in this paper replaces two of the components in the diagram: the fixed image and the metric. This is illustrated in Figure 4, where the two input objects are now a fixed narrow-band and a moving image. The narrow-band acts as a shape representation model of an anatomical structure and can be extracted using a segmentation method. Note that for convenience, the narrow-band is the fixed or target object in the registration process so that the narrow-band does not need to be regenerated for each evaluation of the metric. In many applications, registering the shape model to the image is desired. This is still possible in our scheme as long as the transform used is invertible. At the end of the registration process, all that is required is to invert the transformation and resample the binary shape image instead of the moving input image. In order to evaluate the matching between the narrow-band and the moving image, a

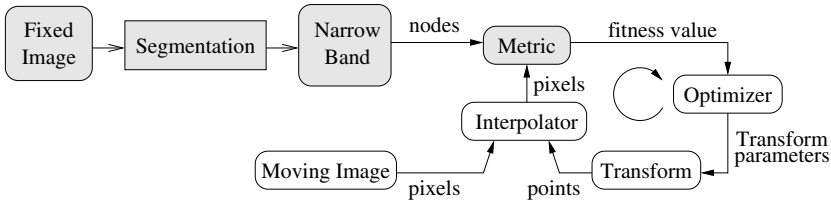


Fig. 4. Modular components involved in a narrow-band to image registration process in ITK.

new metric is introduced. This metric is a fundamental piece of the registration method presented in this paper. Although we analyze a particular metric here, it is conceivable to implement many other metrics in the same framework taking advantage of reusing the rest of modular components. In Figure 4 all the new components are highlighted by shading.

3 Narrow Band to Image Metric

The metric used to evaluate the matching between the narrow-band and the resampled image is the most fundamental component of the registration process. Given that the metric is used as a cost function by the optimizer, its behavior has a significant impact on the performance of the registration process. It determines among others, the computational complexity, capture radius, robustness and stability of the registration.

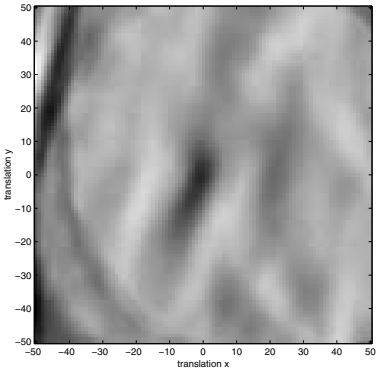


Fig. 5. Grayscale map of the metric values as a function of translation using the brain image and extracted narrow-band from Figure 2.

The metric proposed in this paper is based on the concept of *normalized correlation* [6]. Many other metrics are conceivable. However, for the sake of compactness we restrict ourselves to this particular metric in an initial treatment of the narrow-band to image registration approach.

The metric implemented here works as follows. For each node of the narrow-band, its index position is converted to the coordinates of the physical space. This is done taking into account the origin and spacing of the image from which the narrow-band was extracted. The physical position of the point $\mathbf{p}$ is mapped into the moving image space using the spatial transform $T(\mathbf{p})$. The intensity $I(T(\mathbf{p}))$ of the moving image is resampled at that position $T(\mathbf{p})$. This intensity value is used along with the value from the narrow-band node in order to compute contributions to the correlation terms.

The following equation illustrates the computation performed:

$$NC(A,B) = F \times \frac{\sum_i^N (A_i \cdot B_i)}{\sqrt{\sum_i^N A_i^2 \cdot \sum_i^N B_i^2}} \tag{1}$$

A_i is the value of the i-th node in the narrow-band. B_i is the pixel value of the moving image where the i-th narrow-band node was mapped. N is the number of nodes in the narrow-band. The factor F is set to $+1$ or -1 depending on the gray level characteristic of the anatomical object in the image with respect to its background. For example, when the object is brighter than its background, the factor F is set to $+1$ and the metric is used as a cost function to be minimized by the optimizer. When, on the other hand, the anatomical object happens to be darker than the background, the factor F is set to -1 and again the metric is used as a cost function to be minimized. Changing the F factor is equivalent to inverting the polarity of the narrow-band. That is, making that the values inside the zero set to be positive and the values outside to be negative. This can be interpreted as "matched filtering" of the distribution of intensity levels of the anatomical object in the transition region between the object and the background.

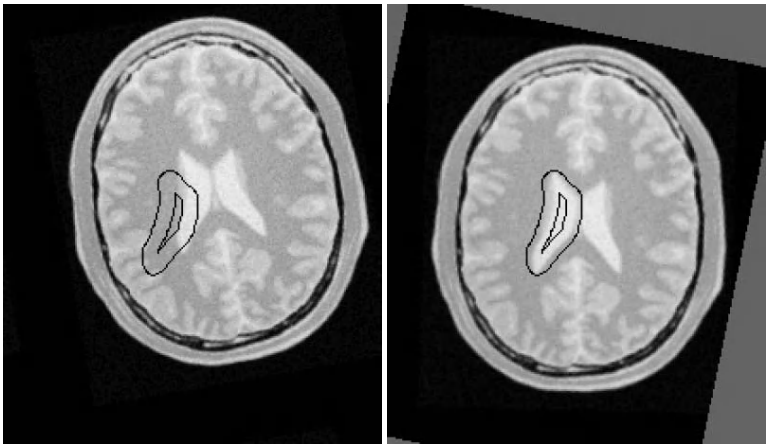


Fig. 6. Superposition of the narrow-band boundaries on the moving image before (left) and after (right) registration was performed. In this case, misalignment due to both rotation and translations is to be corrected.

Characterizing the metric is one of the first tasks to be performed when a new metric is to be introduced. This analysis is presented in the following section.

4 Results

A number of experiments were performed with the aim of characterizing the behavior of this new registration technique when used for medical image applications. The extracted narrow-band in Figure 2 (center) was used as the fixed or target object and various transformed versions of the MRI proton density image of the brain (left) were used as the moving image to be registered onto the target shape.

4.1 Metric Space as Function of Translation

In order to develop an understanding of the metric's behavior, an image of the metric value under various translations of the brain image was computed. This image is presented in Figure 5. The grayscale values represent the metric values ranging from black (lowest value) to white (largest value). Since the narrow-band was originally extracted from the brain image, the two are already in alignment hence the large dark region at the center of the image is associated with zero translation. Notice that the shape of the capture region is similar to that of the segmented ventricle shape and that the region is fairly broad (approximatedly 30 x 60 pixels). There is also another visible minima to the right of center. This is produced by the other ventricle on the right which is also long and thin. However, the metric values in this region have a shallower minima indicating a poor fit, due the other ventricle being approximatedly the mirror image of the template shape.

4.2 Rigid Registration in 2D

In this experiment, we test out the registration process as illustrated in Figure 4. In the example shown in Figure 6, we have used a rotated (10 degrees) and translated ($13mm$ horizontally, $17mm$ vertically) version of the brain image as the moving image. A linear interpolator was used to resample the moving image at non-grid positions. We used the normalized correlation metric as the cost function and connected it to a steepest descent optimizer modified for using an adaptive step [10]. Figure 6 presents the overlay of the

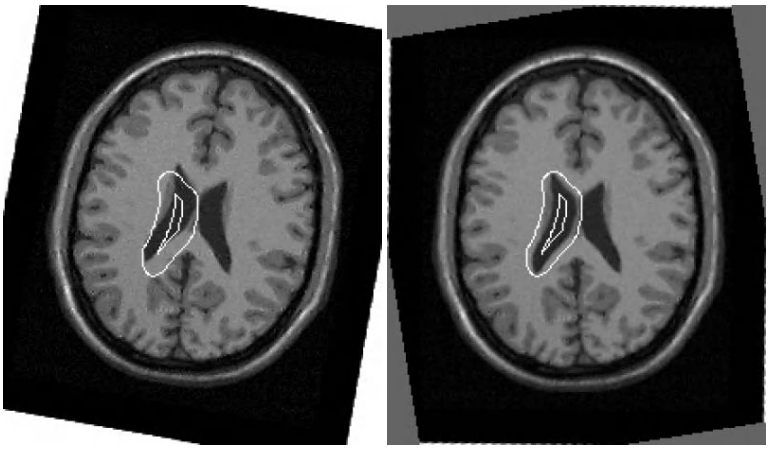


Fig. 7. Multi-modality registration. Superposition of the narrow-band boundaries onto the moving image before (left) and after the registration (right) was performed. In this case, only misalignment due to rotation is to be corrected.

narrow-band boundaries on top of the moving image before and after the registration process. Note that the narrow-band is fixed in space. The image on the right is the moving image resampled using the transform obtained from the registration process. From the image, it can be seen that the registration has corrected the introduced rotation and translation. On a Pentium 4 machine (2.4GHz, 2GB RAM) the registration process took 0.954 seconds. In comparison, a full image to image registration using normalized correlation of the image intensity took 23.891 seconds.

4.3 Multi-modality Registration and Polarity Inversion

As described in section 3, there are cases in which the polarity of the narrow-band must be inverted depending on the intensity profile of the moving image. In the previous example (Figure 6), the polarity of the narrow band was negative inside as illustrated in Figure 2. In the current example we address the reverse case, where the narrow-band is positive inside in order to produce negative correlations with a T1 MRI image. In T1 modality, the ventricles appear as a dark object against a bright background. Figure 7 shows the behavior of the narrow-band to image registration in a multi-modality case.

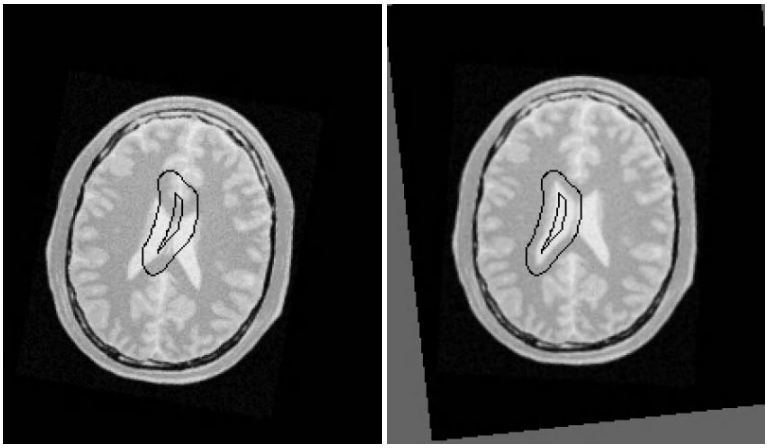


Fig. 8. Illustration of the narrow-band tolerance for small scale changes in the image. The moving image has been reduced by a $1 : 1.2$ factor but still rigid registration with the narrow-band succeeds in locating the left ventricle.

The narrow-band was originally extracted from an MRI proton density image of the brain. The moving image is an MRI T1 image intentionally rotated by 10 degrees. The Figure presents the overlay of the narrow-band boundaries on top of the T1 image before registration and after registration. This simple example shows that we can effectively handle multi-modality registration by performing an intermediate narrow-band extraction step in order to model structures of interest.

4.4 Tolerance to Scale Changes

The fact that the narrow-band has a wide support around the edge, furnishes it with a natural tolerance for small scale differences between the model and the shape to be detected. In fact the narrow band, being a distance map has a local scale invariance property. That is, the narrow-band of a slightly scaled version of the shape is quite similar to the narrow-band of the original shape. This property is illustrated in Figure 8 where the moving image has been generated from the brain image, by applying a rotation of 10 degrees, a translation of $13mm$ along x axis and $17mm$ along y axis, and reducing its scale by a $1 : 1.2$ factor. The width of the narrow-band is such that the edges of the scale reduced object are still inside the capture radius of the narrow-band to image metric. The local scale invariance of the narrow-band means that the border of the band still provides enough information to allow alignment of the shape model to the image.

5 Discussion

In this paper, we introduced a novel concept of narrow-band to image registration. We used the narrow-banding techinque of level set methods to compactly represent the

contour of an object. This new technique has several advantages: arbitrarily complex shapes can be modelled and a great reduction in computational speed compared to full image to image registration. The signed distance representation results in a large capture radius in the optimization space as well as tolerance to scale difference between the shape model and the object in the image. The intermediate narrow-band extraction process also allows images of different modalities to be registered using relatively simple computations.

We have also demonstrated the great advantage of using the ITK modular registration framework. The major burden in implementation of our new technique is the narrow-band to image metric, while relying on existing predefined components to provide the rest of the required functionality. Although we have only analyzed one particular metric in this paper, it is conceivable that there exist other metrics with different desirable properties. Using ITK, the testing of new algorithms is simply a matter of swapping out components of the registration process.

Acknowledgements

The images used in the experiments in this paper were taken from the BrainWeb project (http://www.bic.mni.mcgill.ca/brainweb). This work was partially supported by NIH SBIR grant (5 R44 MH062881-03), NIH contracts (N01-LM-9-3532, N01-LM-0-3504) and by the NLM Insight Toolkit (ITK) project.

References

1. Malladi, R., Sethian, J.A., Vermuri, B.C.: Shape modeling with front propagation: A level set approach. IEEE Trans. on Pattern Analysis and Machine Intelligence **17** (1995) 158–174
2. V. Caselles, R.K., Sapiro, G.: Geodesic active contours. International Journal on Computer Vision **22** (1997) 61–97
3. Chan, T., Vese, L.: Active contours without edges. IEEE Transactions on Image Processing **10** (2001) 266–277
4. Sethian, J.: Level Set Methods and Fast Marching Methods. Cambridge University Press (1996)
5. Maintz, J.B.A., Viergever, M.A.: A survey of medical image registration. Medical Image Analysis **2** (1998) 1–36
6. Penney, G.P., Weese, J., Little, J.A., Desmedt, P., Hill, D.L.G., Hawkes, D.J.: A comparision of similarity measures for use in 2d-3d medical image registration. IEEE Transactions on Medical Imaging **17** (1998) 586–595
7. Borgefors, G.: Hierarchical chamfer matching: A parametric edge matching algorithm. IEEE Transactions on Pattern Analysis and Machine Intelligence **10** (1988) 849–865
8. Iverson, L.A., Zucker, S.W.: Logical/Linear Operators for Images Curves. IEEE Transactions on Pattern Analysis and Machine Intelligence **17** (1995) 982–996
9. Marr, D.: Vision. W. H. Freeman and Company, San Francisco, CA (1982)
10. Ibanez, L., Schroeder, W.: The ITK Software Guide. Insight Consortium, http://www.itk.org/ItkSoftwareGuide.pdf. (2003)

Non-rigid Registration of Mammograms Obtained with Variable Breast Compression: A Phantom Study

Frédéric J.P. Richard[1], Predrag R. Bakić[2], and Andrew D.A. Maidment[2]

[1] MAP5, FRE CNRS 2428, University Paris 5 - René Descartes
Department of Mathematics and Computer Science
45, rue des Saints Pères, 75270 Paris Cedex 06, France
richard@math-info.univ-paris5.fr
[2] Hospital of the University of Pennsylvania
Department of Radiology, 1 Silverstein Bldg
3400 Spruce St., Philadelphia, PA 19104, USA
{bakic,maidment}@rad.upenn.edu

Abstract. The amount of breast compression applied during a mammographic exam affects the appearance of mammograms by introducing variations in the shape, position, and contrast of breast anatomical structures, which can conceal existing breast abnormalities or generate false alarms. Due to the complex tissue organization and elastic properties of the breast and the projective nature of mammography, rigid registration approaches are not useful in correcting these variations. We describe a non-rigid approach focused on registration of mammogram regions of interest, taking into account the changes in image contrast. This registration algorithm has been applied to synthetic mammograms generated using a deformable 3D anthropomorphic phantom and a model of breast deformation during mammographic compression.

Keywords: image registration, partial differential equations, multigrid optimization, finite elements, multimodality registration, mammography, breast compression, tissue modeling, mammogram synthesis.

1 Introduction

Image registration has been an active topic of research for over a decade (see [1] for a recent survey). The most common medical application is in brain imaging [2, 3]. Although less studied, mammogram registration is not only a challenging problem but also an important issue for computer-aided diagnosis. One of the main approaches to tumor detection consists of locally comparing mammograms and identifying abnormal differences [4–8]. This approach can be applied either to successive mammograms of the same breast or mammograms of both the left and right breasts. In the first case, abnormal differences are a sign of possible lesion growth, whereas in the second case they are a sign of suspicious asymmetry. Unfortunately, such comparisons lack specificity due to the large number of normal mammogram differences which are locally similar to abnormalities and cause false-positives.

The main problem in the design of a robust comparison technique is to reduce false-positives by recognizing normal mammogram differences. Normal differences can be

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 281–290, 2003.

due to several factors: acquisition differences, breast positioning and compression variations, and anatomical or histological variations. Differences resulting from acquisition conditions are often very prominent in successive mammograms. They can be corrected by mammogram grey-scale normalization [9]. Differences due to breast positioning can be easily corrected by alignment procedures which involve rotation and translation computed using breast contours [8]. But, the correction of differences due to the last three factors (compression and histological and anatomical variations) is still an open problem. The effect of these factors on a mammogram are not well-known. Modeling mammographic compression effects is an important recent topic of research [10, 11, 9, 12]. Non-rigid image registration techniques have been proposed to correct image differences that remain after mammogram normalization and alignment [4, 5, 7]. However, none of these works has systematically evaluated the ability of techniques to correct each type of difference. Such evaluations are difficult to implement because they require very specific mammogram databases that are not available in the public domain.

Our motivation is to automatically correct mammogram differences due to breast compression variations. Our main contribution is two fold. First, based upon recent works of F. Richard and L. Cohen [4, 5], we propose a new image registration technique which enables such corrections. Second, we show preliminary results of an evaluation based upon synthetic mammograms simulated with a deformable breast phantom [12, 13]. Here we present a report of our work in progress.

In [4, 5], F. Richard and L. Cohen proposed an image-matching approach that focuses on regions of interest. This approach allows a combination of constraints which are intensity and contour based. Such a combination is well-suited for mammogram registration. Breast contours are the most widely used constraint for registration, as they provide robust corrections of breast shape differences. However, these constraints are not sufficient to register accurately changes in internal breast anatomy. Such registrations can be obtained by completing models with some other constraints based on image intensity. In [4, 5], intensity-based constraints are defined using the mean square distance (MSD) between image grey levels (see section 2.1). This similarity measure is generally used in registration of images from a single modality and is well-suited for cases where grey levels are approximately the same from one image to another. Large variations of breast compression can, however, significantly change image contrast. In order to overcome this limitation, we have modified the registration approach allowing linear change of image contrast, similarly to the registration of multimodality images [14, 15]. This modified approach, however, differs by enabling non-rigid deformations, while still combining intensity-based with contour-based constraints. In Section 2, we describe this new registration approach. In Section 3, we present the model which is used to simulate mammograms of a breast with different amounts of compression. In Section 4, we show some registration experiments using simulated mammograms.

2 Registration Technique

2.1 Framework

Let Ω be a connected and open set of $\mathbb{R}^2$ and I^0 and I^1 be two images defined on Ω using interpolation. Let us denote by $\overline{\Omega}$ the set which is the closure of Ω (with

respect to the Euclidean norm of $\mathbb{R}^2$), i.e., the set which contains the set Ω and its boundary. Matching images I^0 and I^1 consists of finding a geometric deformation ϕ such that the deformed image $I^0 \circ \phi$ is "similar" to the target image I^1. Usually [16–18], images are registered on the whole domain Ω and geometric deformations are defined as functions mapping Ω onto itself. As in [4, 5], we focus rather on mapping image regions of interest. For that, we assume that images I^0 and I^1 have single regions of interest which are respectively located on the connected and open subsets Ω_0 and Ω_1 of Ω. We denote by $\partial\Omega_0$ and $\partial\Omega_1$ the boundaries of Ω_0 and Ω_1, respectively. In the registration model defined next, boundaries $\partial\Omega_1$ in I^1 are assumed to be segmented and known whereas boundaries $\partial\Omega_0$ are unknown and segmented during the registration process. We define geometric deformations ϕ on the known region of interest $\overline{\Omega}_1$ of I^1. These deformations are elements of a space $\mathcal{W}$ which is composed of smooth functions mapping the domain $\overline{\Omega}_1$ into $\overline{\Omega}$. We will denote by u a displacement fields associated with deformations ϕ. Displacements u also belong to $\mathcal{W}$. They are equal to $\phi - \mathrm{Id}$, where Id is the identity map of $\mathcal{W}$ (i.e. $\forall\, x \in \overline{\Omega}, \mathrm{Id}(x) = x$).

In registration tasks involving a single modality, the image similarity criterion is usually the MSD. For all images I^0 and I^1 in $L^2(\Omega)$ and any open subset U of Ω, this distance is defined as:

$$D^2_U(I^0, I^1) = |I^0 - I^1|^2_U = \int_U (I^0(x) - I^1(x))^2 \, dx. \tag{1}$$

This distance is low if on each point of U, grey levels values of both images I^0 and I^1 are close. This distance is not invariant with contrast changes. Hence, a MSD criterion is not suitable for the similarity quantification of images having different contrasts. One of the ways to define a criterion $D_{U,\mathcal{C}}$ which is invariant to a group $\mathcal{C}$ of contrast changes is the following:

$$D^2_{U,\mathcal{C}}(I^0, I^1) = \inf_{g \in \mathcal{C}} |g \cdot I^0 - I^1|^2_U. \tag{2}$$

In this definition, $g \cdot I^0$ is the action of a contrast change g of $\mathcal{C}$ on I^0. In this paper, we will simply use a measure which is invariant to linear contrast changes. Hence, in what follows, the set $\mathcal{C}$ will be equal to $\mathbb{R}^2$ and the action of an element $g = (g_1, g_2)$ of $\mathcal{C}$ on an image I will be defined as

$$g \cdot I = g_1 \, I + g_2. \tag{3}$$

In this equation, g_1 is a grey-level dilatation factor and g_2 is a translation factor. Let ϕ be in $\mathcal{W}$ and g in $\mathcal{C}$, we will denote by $I_{\phi,g}$ the geometric deformation of the image I under the contrast change g:

$$\forall\, x \in \Omega_1, \;\; I_{\phi,g}(x) = g \cdot I \circ \phi(x) = g_1 \, I(\phi(x)) + g_2. \tag{4}$$

When $g = (1, 0)$, $I_{\phi,g}$ will be also denoted I_ϕ.

2.2 Mathematical Model

The registration problem is stated in terms of an inverse problem as follows:
Problem 1 (first formulation) *Find an element of $\mathcal{W}$ which minimizes an energy $\tilde{J}$ of the following form:*

$$\tilde{J}(u) = \frac{1}{2}\, A_{\Omega_1}(u,u) + \frac{\gamma_1}{2}\, D^2_{\Omega_1,\mathcal{C}}(I^0_\phi, I^1) + \gamma_2 \int_{\Omega-\phi(\Omega_1)} S((I^0(x))^2)dx, \quad (5)$$

with free boundary conditions on $\partial\Omega_1$. In this energy definition, parameters γ_1 and γ_2 both belong to $\mathbb{R}^+$.

The energy in Equation (5) is composed of three terms. The first term is a smoothing term which ensures that the problem is well-posed and that solutions are non-degenerate. As in [5, 4], the design is based on a strain energy of the linearized elasticity. The second term is the image similarity measure defined in Equation (2). It is an intensity-based registration constraint. The third term is a registration constraint which ensures that energy minima map Ω_1 onto Ω_0. It is defined on a region $\Omega - \phi(\Omega_1)$ which is an expected background of I^0. The function $S(y^2)$ is low when y is likely to be a background grey-level value and high when it is not. Due to the third term, the registration model also enables the segmentation of an otherwise unknown region of interest in I^0. After minimization, boundaries of this region of interest are described by $\phi(\partial\Omega_1)$.

It can be shown using Green's formulae that Problem 1 is equivalent to the following problem [4, 5]:

Problem 1 (equivalent formulation) *Find an element of $\mathcal{W} \times \mathcal{C}$ which minimizes an energy J which is of the following form:*

$$J(u,g) = \frac{1}{2}\, A_{\Omega_1}(u,u) + \frac{\gamma_1}{2}|I^0_{\phi,g} - I^1|^2_{\Omega_1} - \gamma_2 \int_{\Omega_1} S((I^0_\phi(x))^2)\, \det(\nabla\phi)\, dx, \quad (6)$$

with free boundary conditions on $\partial\Omega_1$.

In this equation, the real value $\det(\nabla\phi)$ is the Jacobian of ϕ. In this formulation, contrast changes g are explicitly mentioned as unknown variables.

2.3 Numerical Solution

The gradient $\nabla J_{u,g}$ of the energy J (Equation (6)) with respect to the variable u in $\mathcal{W}$ was computed in [5, 4]. It is given by

$$\nabla J_{u,g} = u - L^{-1}\, f(\phi, g), \quad (7)$$

where L is the operator of the linearized elasticity [5, 4] and f is given by

$$\begin{aligned} f(\phi,g) = &-\gamma_1\, (I^0_{\phi,g} - I^1)\, \nabla I^0_\phi \\ &+2\,\gamma_2\, \det(\nabla\phi)\, S'((I^0_\phi)^2)\, \nabla I^0_\phi - \gamma_2\, \mathrm{div}\{S((I^0_\phi)^2)\, \mathrm{cof}(\nabla\phi)^T\}, \end{aligned} \quad (8)$$

where $\mathrm{cof}(M)$ is the cofactor matrix of a matrix M ($\mathrm{cof}(M) = \det(M)M^{-T}$).

Now, let u be fixed in $\mathcal{W}$, the function G_u which is defined on $\mathbb{R}^2$ and associates g to $J(u,g)$ is convex. Thus, the minimum of G_u is the unique solution of Euler equations $\partial_{g_1} G_u(g) = 0$ and $\partial_{g_2} G_u(g) = 0$, where $\partial_{g_1} G_u(g)$ and $\partial_{g_2} G_u(g)$ are respective partial derivatives of G_u:

$$\partial_{g_1} G_u(g) = \gamma_1 \big(g_1 \int_{\Omega_1} (I^0_\phi(x))^2 dx + g_2 \int_{\Omega_1} I^0_\phi(x) dx - \int_{\Omega_1} I^1(x) I^0_\phi(x) dx \big),$$

$$\partial_{g_2} G_u(g) = \gamma_1 \big(g_1 \int_{\Omega_1} I^0_\phi(x) dx + g_2 |\Omega_1| - \int_{\Omega_1} I^1(x) dx \big),$$

The solution $\hat{g}_u$ of the Euler equations is given by:

$$\hat{g}_{1,u} = \frac{\int_{\Omega_1} I^1(x)\, I^0_\phi(x) dx - \int_{\Omega_1} I^1(x) dx \int_{\Omega_1} I^0_\phi(x) dx}{\int_{\Omega_1} (I^0_\phi(x))^2 dx - (\int_{\Omega_1} I^0_\phi(x) dx)^2}, \tag{9}$$

$$\hat{g}_{2,u} = \frac{1}{|\Omega_1|} (\int_{\Omega_1} I^1(x) dx - \hat{g}_1 \int_{\Omega_1} I^0_\phi(x) dx). \tag{10}$$

Using these remarks, we can derive a gradient descent algorithm for the numerical resolution of Problem 1.

Algorithm 1 (gradient descent)

$$\forall\, t > 0, \;\; \frac{du}{dt}(t) = -u(t) + \delta(t) \quad \text{and} \quad u(0) = M_0, \tag{11}$$

where at each time t, $\delta(t)$ is the solution of the following partial differential equation:

$$L\, \delta = f(\phi(t), \hat{g}(t)), \tag{12}$$

with f defined as in Equation (8) and $\hat{g}(t)$ given by Equations (9) and (10) for $u = u(t)$.

For the implementation of this algorithm, we use the same approach as in [4, 5]: we discretize Equation (12) using the Galerkin method and adopt a multigrid, coarse-to-fine optimization strategy. Also, the initial segmentation of the region of interest in I^0 is computed with the same approach. An initial displacement, M_0, is derived using a coarse matching of regions of interest boundaries.

3 Mammogram Model

Evaluation of the image registration problem was performed using synthetic mammograms based upon an anthropomorphic breast model and a simulation of the mammographic imaging process developed by P. Bakic et al. [12]. The 3D anthropomorphic breast model has been designed with a realistic distribution of large and medium scale tissue structures. Parameters controlling the size and placement of the simulated structures (adipose compartments and ducts) provide a method of consistently modeling images of the same simulated breast with different compression, projection angle and acquisition parameters. The mammographic imaging process is simulated using a compression model and a model of the x-ray image acquisition process. This mammography simulation has been evaluated in terms of mammographic texture [13] and ductal morphology [19], and has been used to estimate the mean glandular dose in mammography [20]. The compression model estimates breast deformation using published values of tissue elasticity parameters and clinically relevant force values. Synthetic mammograms are simulated using a monoenergetic parallel x-ray beam applied to the synthetically compressed breast phantom.

3.1 3D Anthropomorphic Breast Model

The breast model contains two ellipsoidal regions of large scale tissue elements: predominantly adipose tissue (AT) and predominantly fibroglandular tissue (FGT). Analysis of subgross histologic breast images and the corresponding mammograms showed that the background texture, or parenchymal pattern, found in mammograms is predominantly due to the projection of the connective tissue surrounding adipose compartments. These compartments are included in the model to simulate the breast adipose tissue, and these form the medium scale breast model elements. In addition, the model includes a representation of the breast ductal network. The adipose compartments are approximated by thin shells in the AT region and small blobs in the FGT region. The interiors of the shells and blobs have the elastic and x-ray attenuation properties of adipose tissue; while the shell layer and the portion of the FGT region surrounding blobs simulate the properties of the connective tissue. As a first approximation, the adipose compartments are represented by spheres. The size of the spheres can vary to allow for normal breast anatomic variations. depending upon the amount of adipose tissue in the breast. Adipose compartments are more easily identified in a histology image than in a mammogram, since the latter image contains the superimposed projections of many tissue layers. Generation of the simulated adipose compartments is described in more detail in [12] and generation of the binary tree models of the breast ductal networks is given in [19].

3.2 Simulation of Mammographic Compression

Mammographic compression is simulated based upon tissue elasticity properties and a simplified breast deformation model. Deformation is simulated separately for slices of the breast model positioned normal to the compression plates. Each slice is approximated by a beam composed of two different tissues. The deformed slices are stacked to produce a model of the compressed breast.

Tissue elasticity parameters from the literature vary significantly, partly because they have been measured experimentally using small samples taken from a particular tissue type. The breast, however, consists of a complex admixture of different tissues types, which affects the elastic behavior of the whole organ. We used parameters calculated from the speed of sound in tissue [21] and tissue density. The Mammography Quality Standards Act [22] recommends the minimum and maximum compression force to be used in mammography. There are also reports in literature of statistical analyses of the force and compressed breast thickness measured during exams [23]. The values of the force used in the mammographic simulation were selected based upon these two constraints.

In this paper we have generated medio-lateral oblique (MLO) mammographic views calculated with various amounts of compression. The breast compression simulation is performed in the following steps. First, a slice, with thickness equal to the desired image resolution (200 micron/pixel in this experiment) is approximated by a rectangle with total area and nipple-chest wall dimension equal as in the original slice. The FGT region within the slice is approximated by a rectangle which again has the same area and nipple-chest dimension as in the original slice. These approximations are designed

so that the distance between the centers of gravity of the whole slice and the FGT portion is also the same as in the original slice. Next, the rectangular approximation is deformed using Hooke's law and elastic moduli values corresponding to the FGT and the surrounding AT regions. Finally, the deformed rectangular approximation is used to compute the compressed phantom slice shape. Attention is taken to realistically simulate the flattened shape of the compressed breast. As such, the breast thickness is equal to the distance between the compression plates everywhere except in a narrow region close to the anterior edge of the breast. Separate processing of individual model slices is performed, followed by restacking the deformed slices together to form the 3-D compressed breast model. Computation details can be found in [12].

4 Results and Discussion

We have applied the modeling approach described in Section 3 to synthesize mammographic images of the same simulated breast under varying amounts of compression. Examples are shown on Figures 1(a) and (d). Varying breast compression has several effects on the appearance of the mammograms. The size of the breast domain in the mammogram increases as the breast becomes more compressed and the shape changes slightly. As the compression is varied, the FGT region, the central bright region composed of predominantly fibro-glandular tissue, also undergoes significant geometric deformations. As the compression decreases, the center of this region translates in the direction of the chest wall. The FGT region also changes its shape in a way that is indicative of a non-rigid deformation. Regions where fatty tissues are predominant also vary. In particular, in the top (simulated axilla) of images, the fatty tissue region which is between the FGT and the background becomes narrower as the compression decreases.

Compression variations affect the mammogram contrast. Increasing the amount of compression results in reducing the thickness and spreading the breast FGT region, which, if acquisition parameters remain constant, reduces the image latitude. Such a change can for instance be observed in Figure 2 by comparing histograms of mammograms obtained with different amount of compression. Note that the mammogram obtained with a compression of 8cm appears darker (lower pixel intensities) than the others (see Figure 1). This is because that mammogram was simulated with different acquisition parameters, which corresponds to a clinical situation of manually changing the x-ray technique for breasts of different thickness.

The algorithm described in Section 2 was applied to pairs of synthetic images, with the goal of correcting shape and intensity differences between structures in the unregistered mammograms. Parameter values were selected in accordance with our previous works [4, 5]. Computation times were about twenty seconds on a PC Intel Pentium IV 2.2 GH. A registration example is shown on Figure 1. In this example, the source image I^0 (image (a)) is the mammogram obtained with a breast compression of 8cm. It is deformed onto the target image I^1, which is the mammogram obtained with a breast compression of 5cm. Observing images (a), (b) and (d) and images (e) and (f), several significant corrections of geometric differences between mammogram coarse structures are noticeable. First of all, very few differences due to breast size variations remain after registration. These large differences are corrected as the result of a strong dilation

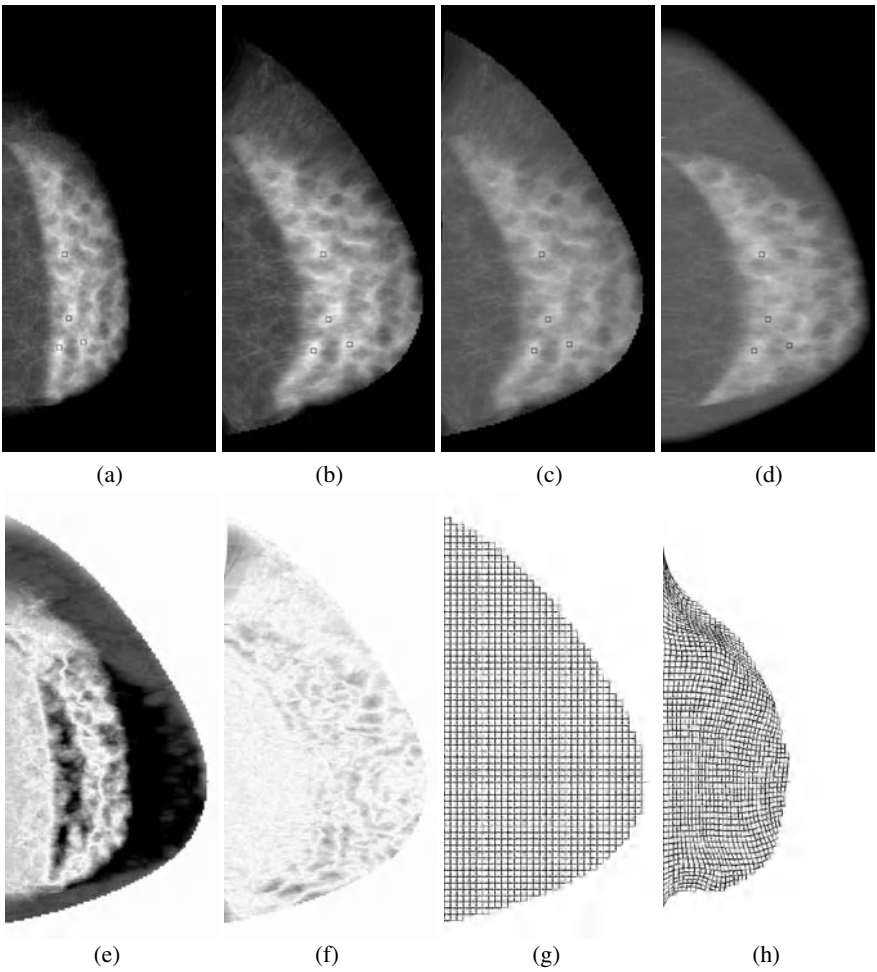


Fig. 1. A mammogram registration example: (a) and (d) are simulated mammograms of the breast with compressions of 8cm and 5cm, respectively. (a) is the source image I^0 and (d) is the target image I^1. (b) is the geometric deformation I^0_ϕ of I^0 that is computed using Algorithm 1. (c) is the contrast corrected deformation $I^0_{\phi,g}$ of I^0. [Small squares in (a), (b), and (d) indicate positions of several corresponding homologous salient points.] Absolute differences between unregistered mammograms (I^0 and I^1) and between registered and contrast corrected mammograms ($I^0_{\phi,g}$ and I^1) are shown respectively in (e) and (f) [black=high differences, white=low differences]. (g) shows the tessellated breast domain Ω_1 of I^1 and (h) the deformed domain ($\phi(\Omega_1)$).

of the source image. Initial differences due to variations in the shape of the FGT region are also significantly reduced. Left FGT boundaries in the source mammogram are particularly deformed. Their shape, which is approximately a straight line, is transformed into a curved line. In addition, in the upper portions of Figures 1(a) and (b), we can observe a large dilation of the source image which corrects shape differences of fatty tissue areas.

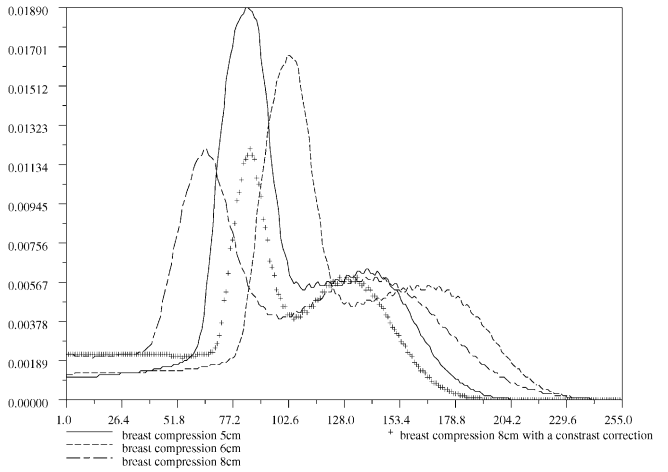


Fig. 2. Comparison of grey-level histograms of simulated mammograms.

The algorithm not only corrects coarse structure shapes. As can be seen in Figure 1(f), it also corrects some differences inside these structures. Such corrections can also be illustrated using salient points. In Figures 1(c) and (d), we selected manually four pairs of homologous salient points having the same positions and computed back their positions in Figure 1(a). The observed displacements of these points (marked with small squares) outline repositioning of homologous points of inner structures during the registration. The algorithm also provides for relevant correction of image contrast variations, as illustrated by Figure 1(c). This correction can also be seen in the comparison of mammogram histograms on Figure 2. As a result of the contrast change, histogram of the mammogram with a breast thickness of 8cm matches the histogram of the mammogram of thickness 5cm.

5 Conclusions

We have described a non-rigid mammogram registration algorithm, modified to provide corrections for image contrast variations. Use of the proposed algorithm was demonstrated on synthetic mammograms generated by applying different amounts of mammographic compression to the same 3D software breast phantom. Further studies are needed, aimed at more precise statistical evaluation of the algorithm.

References

1. J. Maintz and M.A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.
2. U. Grenander and M.I. Miller. Computational Anatomy: an emerging discipline. *Quarterly of Applied Mathematics*, 56(4):617–694, December 1998.
3. A. Toga. *Brain mapping*. Academic Press, 1999.

4. F. Richard and L. Cohen. A new image registration technique with free boundary constraints: application to mammography. In A. Heyden et al., editor, *Proc. ECCV*, volume 4, pages 531–545, Copenhaguen, Denmark, may 2002. Springer.
5. F. Richard and L. Cohen. Non-rigid image registration with free boundary constraints: application to mammography. *Journal of Computer Vision and Image Understanding, Special issue on Nonrigid registration*, 89(2-3):166–196, 2003.
6. F. Richard and C. Graffigne. An image-matching model for the registration of temporal or bilateral mammogram pairs. In M. Yaffe, editor, *Proc. of the 5th International Workshop on Digital Mammography*, pages 756–762, Toronto, Canada, June 2000. Medical Physics.
7. M.Y. Sallam and K. Bowyer. Registration and difference analysis of corresponding mammogram images. *Medical Image Analysis*, 3(2):103–118, 1999.
8. J.L. Semmlow, A. Shadagopappan, L.V. Ackerman, et al. A fully automated system for screening Xeromammograms. *Computers and Biomedical Research*, 13:350–362, 1980.
9. R.P. Highnam and J.M. Brady. *Mammographic image analysis*. Kluwer series in Medical Imaging. Kluwer Academic Publishers, Dordrecht; Boston, March 1999.
10. R. Novak. *Transformation of the female breast during compression at mammography with special reference to importance for localization of a lesion.* PhD thesis, Departement of Diagnostic Radiology at Lakarhuset and Karolinska Sjukhuset, Sweden, 1989.
11. P. Bakic, D. Brzakovic, P. Brzakovic, et al. An approach to using a generalized breast model to segment digital mammograms. In *Proc. of 11th IEEE Symposium on Computer-Based Medical Systems*, pages 84–89, Lubbock, 1998.
12. P. Bakic, M. Albert, D. Brzakovic, and A. Maidment. Mammogram synthesis using a 3D simulation. i. breast tissue model and image acquisition simulation. *Medical Physics*, 29(9):2131–2139, 2002.
13. P. Bakic, M. Albert, D. Brzakovic, and A. Maidment. Mammogram synthesis using a 3D simulation. ii. evaluation of synthetic mammogram texture. *Medical Physics*, 29(9):2140–2151, 2002.
14. F. Maes, A. Collignon, D. Vandermeulen, et al. Multi-modality volume registration by maximization of mutual information. *IEEE Trans. on Medical Imaging*, 16(2):187–198, 1997.
15. P. Viola and W. Wells. Alignment by maximization of mutual information. *Intern. J. of Comp. Vision*, 24(2):137–154, 1997.
16. Y. Amit. A non-linear variational problem for image matching. *SIAM Journal on Scientific Computing*, 15(1):207–224, January 1994.
17. R. Bajcsy and S. Kovacic. Multiresolution elastic matching. *CVGIP*, 46:1–21, 1989.
18. G.E. Christensen, R.D. Rabbitt, and M.I. Miller. Mathematical textbook of deformable neuroanatomies. In *Proc. Natl. Acad. Sci., USA*, volume 90, pages 11944–11948, dec 93.
19. P. Bakic, M. Albert, and A. Maidment. Classification of galactograms using ramification matrices: preliminary results. *Academic Radiology*, 10(2):198–204, 2003.
20. R. Hunt, D. Dance, P. Bakic, et al. Monte carlo simulation of x-ray mammography using a realistic voxel phantom. In *Proc. of UK Radiological Congress*, Birmingham, UK, 2003.
21. G. Kossoff, E.K. Fry, and J. Jellins. Average velocity of ultrasound in the human female breast. *J. Acoust. Soc. Amer.*, 53:1730–1736, 1973.
22. ACR Committee on Quality Assurance in Mammography. *Mammography Quality Control Manual*. Amercan College of radiology, Reston, VA, 1999.
23. D. Sullivan, C. Beam, S. Goodman, and D. Watt. Measurement of force applied during mammography. *Radiology*, 181:355–357, 1991.

An Alternating-Constraints Algorithm for Volume-Preserving Non-rigid Registration of Contrast-Enhanced MR Breast Images

Torsten Rohlfing[1], Calvin R. Maurer, Jr.[1],
David A. Bluemke[2], and Michael A. Jacobs[2]

[1] Image Guidance Laboratories, Department of Neurosurgery
Stanford University, Stanford, CA 94305, USA
rohlfing@stanford.edu
calvin.maurer@igl.stanford.edu
[2] The Russell H. Morgan Department of Radiology and Radiological Science
The Johns Hopkins University School of Medicine
Baltimore, MD 21205, USA
dbluemke@jhmi.edu, mikej@mri.jhu.edu

Abstract. We propose and evaluate in this work a novel optimization strategy for intensity-based non-rigid image registration of contrast-enhanced images with a volume-preservation constraint. Since patient motion correction and volume preservation are to some extent mutually exclusive goals, one is usually faced with a trade-off between volume preservation of contrast-enhancing structures and artifact reduction. We address this problem by repeatedly applying registration passes with alternating incompressibility constraint weights. The novel optimization method alternates between under-constrained registration (allowing the elimination of motion artifacts in the subtraction images) and over-constrained registration (enforcing volume preservation of contrast-enhancing structures). We apply our method to pre- and post-contrast MR breast images from 17 patients. We evaluate our method and compare it to unconstrained and fixed constraint non-rigid registration by blinded visual assessment of maximum intensity projections of subtraction images. The alternating-constraints algorithm was judged to reduce artifacts better then the fixed-constraint algorithm in 11 out of 17 patients and equally well in the remaining 6. The results of this study show the capability of our method to achieve volume preservation and at the same time reduce artifacts very similar to what can be achieved by unconstrained non-rigid registration.

1 Introduction

Non-rigid image registration algorithms based on free-form deformations have recently been shown to be a valuable tool in various biomedical image processing applications. One application of particular clinical interest is the registration of images acquired before and after contrast administration with the purpose

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 291–300, 2003.

of reducing motion artifacts. A major problem with existing algorithms is that when they are applied to pre- and post-contrast image pairs, they often produce transformations that substantially change (generally decrease, but sometimes increase) the volume of contrast-enhancing structures. Tanner *et al.* [1] documented this phenomenon contrast-enhanced magnetic resonance (MR) breast images. We observed the same behavior for lesions in MR breast images [2], as well as for contrast-enhancing vessels in three-dimensional (3-D) digital subtraction angiography using X-ray computed tomography head-and-neck images [3]. Contrast enhancement is an intensity inconsistency between the two images, which is what intensity-based registration algorithms are designed to minimize. This problem severely affects the usefulness of the resulting transformation for volumetric analysis, image subtraction, multi-spectral classification, and pharmacokinetic modeling.

To address this problem, an additional "energy" term is typically added to the intensity-based similarity measure to constrain the deformation to be smooth. We recently introduced a novel incompressibility (local volume preservation) constraint based on the Jacobian determinant of the deformation [2]. Soft tissue in the human body is generally incompressible for small deformations and short time periods. That is, the tissue can be deformed locally, but just like a gelatin-filled balloon, the volume (local and total) remains approximately constant. When computing non-rigid coordinate transformations between pre- and post-contrast images, this knowledge can be incorporated into the registration process by penalizing deviations of the local Jacobian determinant of the deformation from unity.

With an incompressibility constraint, volume preservation and motion artifact reduction are somewhat mutually exclusive goals of the registration. Therefore, there is usually a trade-off when adjusting the constraint weight: high weights lead to good volume preservation but prevent elimination of artifacts, while low constraint weights allow for good artifact reduction at the cost of increased volume loss of contrast-enhancing lesions. We hypothesize the existence of an approximately incompressible non-rigid coordinate transformation that correctly describes patient motion between the two acquisitions. In other words, we assume that motion artifact can in principle be eliminated while preserving the volume of contrast-enhancing lesions. We furthermore hypothesize that residual motion artifacts observed after non-rigid registration using an incompressibility constraint are an effect of the optimization algorithm getting stuck in a local minimum, which it cannot escape from without violating, at least temporarily, the volume preservation condition.

In this paper, we propose a new optimization strategy for intensity-based non-rigid image registration with an incompressibility constraint. Instead of simply constraining the free-form deformation with a weighted incompressibility constraint, we apply a scheme similar to simulated annealing. The underlying rationale is to first under-constrain (by using a smaller incompressibility constraint weight) the deformation, allowing the non-rigid registration to remove motion artifacts at the cost of shrinking contrast-enhancing lesions. In a subse-

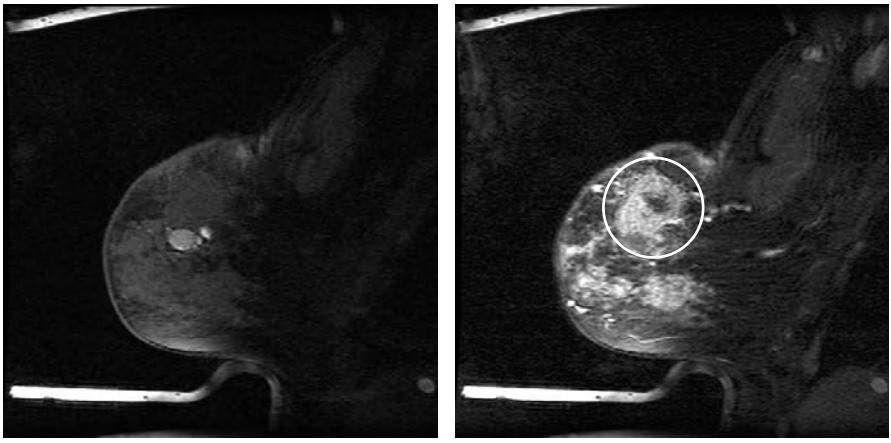


Fig. 1. Image data from one of the patients used in this study. *Left:* Pre-contrast image; *right:* corresponding post-contrast image. The contrast-enhancing lesion is marked with a circle in the post-contrast image.

quent step, the deformation is then over-constrained (by using a larger weight), re-expanding the previously shrunk regions (typically contrast-enhancing lesions) while de-emphasizing the image similarity measure. The purpose of this alternating-constraints strategy is to help the registration avoid local minima by allowing temporary volume loss that is compensated for in a subsequent step after the motion artifacts have been successfully eliminated. Our method is somewhat similar in concept to the alternation of deformation and Gaussian smoothing as described in Ref. [4]. Note, however, that the "smoothing" step in our case is not a purely mathematical operation as it is in Ref. [4], but instead is another deformation step with an increased weight on the regularization term.

To evaluate the efficacy and robustness of our alternating volume-preservation constraint algorithm and in order to validate its clinical value, we apply it to pre- and post-contrast MR breast images from 17 patients. Coordinate transformations were computed by four different algorithms: rigid registration, unconstrained non-rigid registration, fixed constraint non-rigid registration, and alternating constraints non-rigid registration. For all registrations, subtraction images were generated for the registered images and rendered by maximum intensity projection (MIP). The MIP renderings were then randomized and assessed for residual motion artifacts by an expert observer in a blinded evaluation study.

2 Materials and Methods

2.1 Image Data

We applied the non-rigid registration algorithm presented in this paper to MR breast images acquired before and after contrast injection. Seventeen patients

(age range, 18–80 years; median, 45 years) were consecutively referred for MR evaluation of suspicious findings and/or extent of the breast lesion. Each pair of pre- and post-contrast images was first registered using a rigid transformation. Given the rigid registration as the initial alignment, various non-rigid registrations were then computed.

All MR scans were performed on a 1.5T MR scanner (General Electric Medical Systems, Milwaukee, WI), using a dedicated phased array breast coil (MRI Devices, Waukesha, WI) with the patient lying prone with the breast in a holder to reduce motion. Fat suppressed 3-D T_1-weighted FSPGR (T_R/T_E = 20/4 ms, FOV = 18×18 cm, matrix = 512×160, 60 slices, slice thickness 2 mm) pre- and post-contrast images were obtained after intravenous administration of Gd-DTPA contrast agent (Magnevist, Berlex, Wayne, NJ; patients received 0.1 mmol/kg as 0.2 mL/kg of 0.5 mol/l contrast solution). The contrast agent was hand injected over 10 seconds with MR imaging beginning immediately after completion of the injection. The contrast bolus was then followed by a 20 cc flush. The breast lesion was defined by contrast enhancement and identified by a radiologist. A sample pre- and post-contrast image pair is shown in Fig. 1.

2.2 Image Registration Algorithm

An initial alignment of pre- and post-contrast images is achieved using a rigid registration method with six parameters (three for rotation; three for translation). Our algorithm is an independent implementation of a technique for rigid and affine registration described in Ref. [5]. It uses normalized mutual information (NMI) as the image similarity measure [6]. In the first step, this method is used to find an initial rigid transformation to capture the global motion of the object. The rigid transformation is then used as the initial estimate for the non-rigid registration.

The intensity-based non-rigid registration algorithm is an independent, parallel implementation [7] of the technique introduced by Rueckert et al. [8]. The transformation model is a multilevel formulation of a free-form deformation (FFD) based on cubic B-splines. In addition to the NMI similarity measure E_{NMI}, our technique incorporates a penalty term $E_{\mathrm{constraint}}$ to constrain the deformation of the coordinate space [2]. A user-defined weighting factor w ($0 \leq w \leq 1$) controls the relative influence of E_{NMI} and $E_{\mathrm{constraint}}$, combining both into the overall cost function E_{Total} as follows:

$$E_{\mathrm{Total}} = (1 - w)E_{\mathrm{NMI}} - wE_{\mathrm{constraint}}. \tag{1}$$

The design of our deformation constraint is motivated by the observation that most tissues in the human body, including the breast, are approximately incompressible for small deformations and short time periods. In a small neighborhood of the point (x, y, z), the local compression or expansion caused by the deformation can be calculated by means of its Jacobian determinant. The value of the Jacobian determinant is equal to 1 if the deformation at (x, y, z) is incompressible, greater than 1 if there is local expansion, and less than 1 if there is

compression. The incompressibility constraint penalty term we use is the integral of the absolute logarithm of the Jacobian determinant, integrated over the domain of the reference image [2]. This term penalizes local deviations of the Jacobian determinant from unity, that is, it penalizes local tissue expansion and compression. Alternative forms of the incompressibility constraint term are possible [2].

Like Rueckert et al. [8], we aim to improve robustness and efficiency of the registration algorithm by employing a multiresolution approach, starting with a coarse control point spacing that is successively refined using a B-spline subdivision algorithm. The alternating-constraints strategy is applied at each level, starting with an under-constrained optimization pass[1], followed by an over-constrained pass. A B-spline subdivision, if the finest control point resolution has not yet been reached, takes places after an over-constrained pass. The constraint weight for the over-constrained pass was adjusted individually for each patient; the respective constraint weight for the under-constrained pass was defined as 1/100 of the over-constrained weight.

2.3 Study Design

We evaluate and compare the artifact reduction achieved by four different registration methods: rigid, unconstrained non-rigid, non-rigid with a fixed constraint weight, and non-rigid with alternating constraint weights. For this study, multiple constrained free-form deformations were computed, covering a large range of weighting factors for the the deformation constraint. For the fixed-constraint algorithm the weights ranged from 0.0001 to 0.5, while for the alternating-constraints algorithm the weights ranged from 0.01 to 1. The latter values refer to the over-constrained pass; the weights in the under-constrained pass were determined by multiplying the over-constraining weights by 0.01. All non-rigid registrations started with a 40 mm control point spacing that was successively refined to 20 mm, 10 mm, and finally 5 mm. In parallel, the image data resolution was refined from 4 mm voxel size to 2 mm, 1 mm, until at the final stage the original image data was used. Table 1 gives an overview of the parameters used at each level for fixed-constraint and alternating-constraints non-rigid registration.

A key question to answer is whether a given technique provides a way of compensating motion artifacts while preventing volume loss of contrast-enhancing structures. For each algorithm (fixed and alternating constraints) and each patient, we independently determined the smallest weighting factor for the respective algorithm that provided volume preservation of the identified lesion within 2% of the original volume. As reference methods, rigid and unconstrained non-rigid transformations were also computed. For the resulting four transformations, 3-D subtraction images of the registered pre- and post-contrast images

[1] The term "optimization pass" in this context refers to a complete optimization of the non-rigid transformation parameters at one particular voxel size and control point resolution.

Table 1. Algorithm parameters during multi-level non-rigid registration.

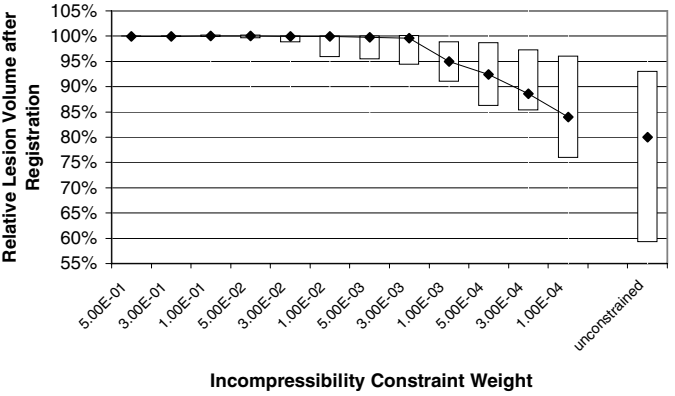
Algorithm	Level	CPG Spacing	Voxel Size	Constraint Weight
Fixed Constraint	1	40 mm	4 mm	w
	2	20 mm	2 mm	w
	3	10 mm	1 mm	w
	4	5 mm	original	w
Alternating Constraints	1	40 mm	4 mm	$10^{-2}w$
	2	40 mm	4 mm	w
	3	20 mm	2 mm	$10^{-2}w$
	4	20 mm	2 mm	w
	5	10 mm	1 mm	$10^{-2}w$
	6	10 mm	1 mm	w
	7	5 mm	original	$10^{-2}w$
	8	5 mm	original	w

were computed, rendered by lateral orthogonal maximum intensity projection[2] (MIP), randomized, and presented to an expert observer. The blinded observer ranked the four subtraction MIP images based on motion artifact reduction. Rank #1 was assigned to the image showing least residual artifact, while rank #4 was assigned to the image showing most residual artifact. Several images could be assigned the same rank, in which case an appropriate number of subsequent ranks would be left unassigned. For example, if one image clearly showed least artifact, another clearly showed most artifact, and the remaining two images were indistinguishable, then the ranks assigned would be #1, #2, #2, and #4 with rank #3 unassigned.

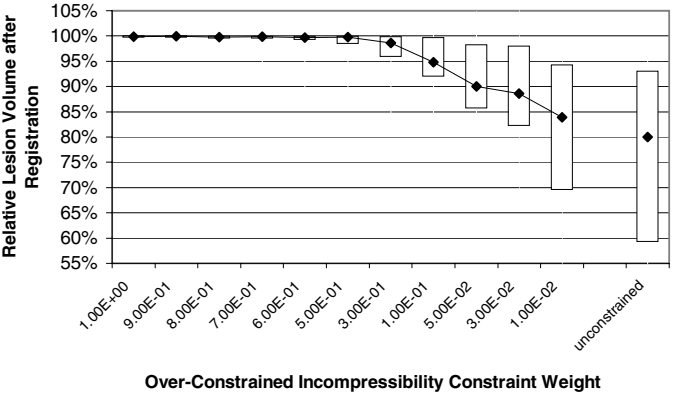
3 Results

Volume Preservation. The original volumes of the contrast-enhancing lesions in the images from 17 patients ranged from 0.2 to 77.7 ml (mean ± SD = 9.1 ± 19.0 ml). These volumes were determined by semi-automatically segmenting the contrast-enhancing lesions in the subtraction image after rigid registration. As other groups have reported before, unconstrained non-rigid registration of pre- and post-contrast images often results in substantial volume loss of the contrast-enhancing structures. For the 17 patients considered in the present study, volume loss was between 1.3 and 61.8 percent (mean ± SD = 23.4 ± 18.3 percent). On average, for both optimization strategies (fixed and alternating constraints), volume preservation improves as the relative weight of the constraint penalty term increases (Fig. 2). With sufficiently high relative weighting factors, the median volume change of contrast-enhancing lesions is less than 1% for both strategies.

[2] Since the original plane orientation of the MR images used in this study was sagittal, a lateral projection showed the most image detail and was therefore considered most appropriate for image quality assessment.



(a) Fixed Constraint Weight



(b) Alternating Constraint Weights

Fig. 2. Relative volumes of contrast-enhancing lesions after non-rigid registration with different incompressibility constraint weights.

Artifact Reduction. The results of the expert assessments of the artifact reductions achieved on the 17 patients in this study are shown in Fig. 3. For each constrained non-rigid registration optimization strategy and each patient, we used the deformation produced by the smallest weight factor that produced volume preservation of the identified lesion within 2%. Unconstrained non-rigid registration was ranked best (least residual motion artifact) in all cases. The alternating-constraints algorithm clearly outperformed the fixed constraint algorithm: in 11 out of 17 cases, the alternating-constraints method was ranked better than the fixed-constraint method, and equally good in the remaining 6 cases. Paired two-sided t-tests on the distribution of ranks showed that the differences among all algorithms are statistically significant. The significance level

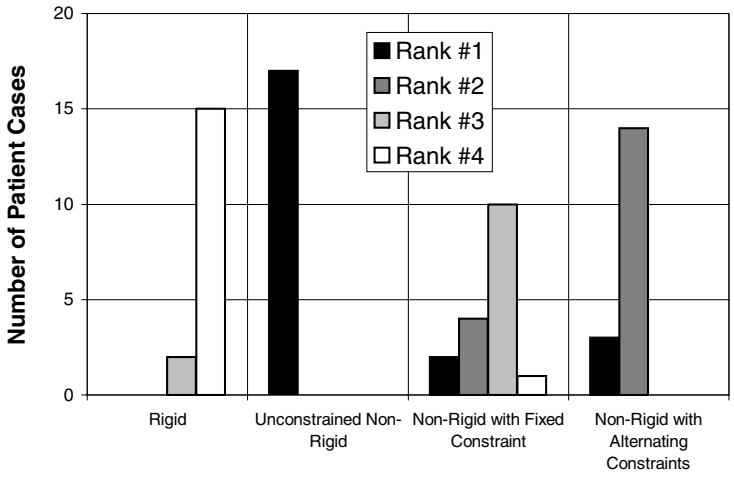


Fig. 3. Expert assessment of residual artifacts in subtraction MIP renderings after motion correction by four different registration methods. The bars show the number of patients for which each quality rank was assigned to the respective registration method (least artifact, Rank #1; most artifact, Rank #4).

for the superiority of the alternating-constraints approach over fixed-constraint registration is $P < 2.2 \cdot 10^{-4}$.

Computation Times. For the fixed-constraint algorithm, the median computation time over 204 registrations (17 patients, 12 constraint weights per patient) was 24,344 seconds. For the alternating-constraints algorithm, the median computation time over 187 registrations (11 constraint weights per patient) was 27,406 seconds. The alternating-constraint algorithm, despite running twice as many optimization passes, was only about 12% slower than the fixed-constraint algorithm. We observed that the over-constrained passes typically required substantially fewer optimization steps than the under-constrained passes, thereby explaining the comparatively low additional computational cost.

4 Conclusion

Results on contrast-enhanced MR breast images from 17 patients suggest that the incompressibility constraint improves non-rigid registration of pre- and post-contrast images by substantially reducing the problem of shrinkage of contrast-enhancing structures. Motion artifacts can still be reduced substantially, but the more strictly the constraint is enforced (larger relative weight in the cost function), the more residual artifacts remain.

Using a novel alternating-constraints registration optimization strategy, artifact reduction and lesion volume preservation were both reasonably well achieved simultaneously. In a blinded assessment study, the alternating-constraints algo-

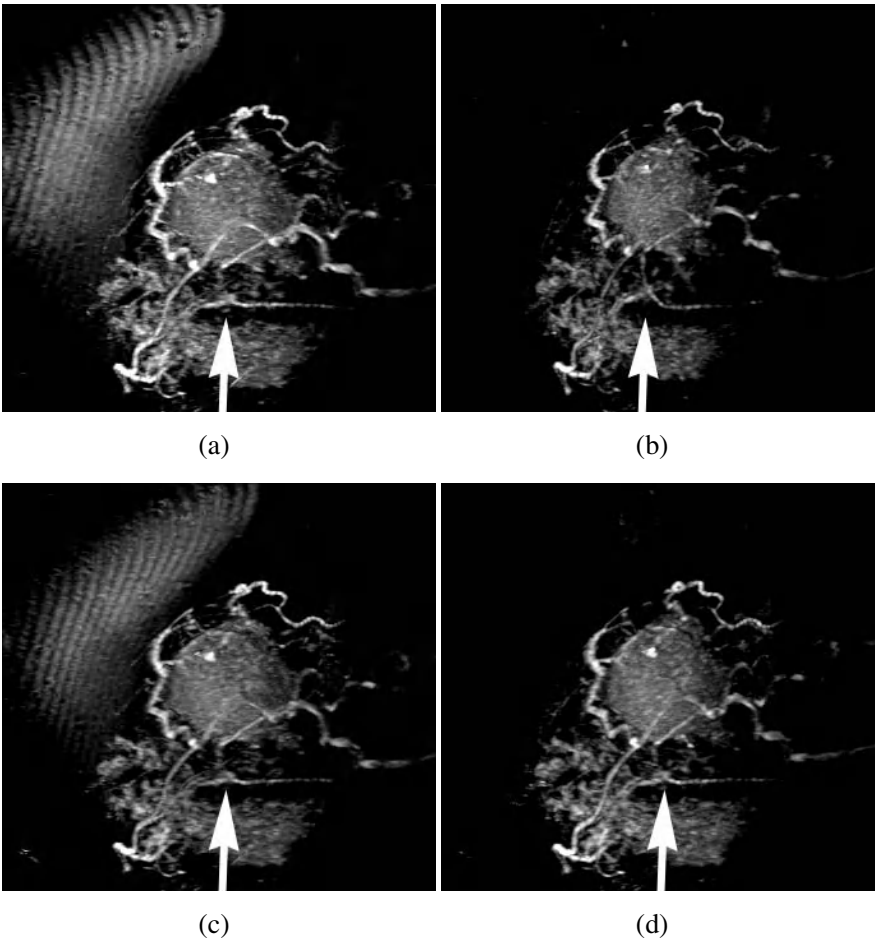


Fig. 4. Effect of alternating constraints on motion artifacts in MIP of subtraction images. (a) After rigid registration. (b) After unconstrained non-rigid registration (30% volume loss). (c) After non-rigid registration with single incompressibility constraint (1% volume loss). (d) After non-rigid registration with alternating incompressibility constraints (1% volume loss). In the blinded evaluation of the MIP renderings from this patient, the expert rated unconstrained non-rigid registration slightly superior (i.e., as having slightly less residual artifacts) to the alternating-constraints algorithm. The fixed-constraint algorithm left substantial residual artifacts and was ranked third, followed by rigid registration. The unconstrained non-rigid registration reduces motion artifacts but shrinks the contrast enhancing lesion by 30% and also causes some change in the shape of structures, e.g., the arrow points to a bending that appears in a contrast-enhancing vessel after unconstrained registration. The incompressibility constraint with the normal search strategy preserves the lesion volume and the vessel morphology but does not in this case compensate for motion artifacts. The incompressibility constraint with alternating constraints optimization preserves the lesion volume and the vessel morphology and also substantially reduces motion artifacts.

rithm clearly outperformed the fixed-constraint algorithm. Although the unconstrained non-rigid registration resulted in the fewest residual artifacts, in many cases the perception of the expert observer was that the differences between the alternating-constraints method and unconstrained non-rigid registration were very subtle or indistinguishable. An example result is illustrated in Figure 4. The additional computational cost incurred by the alternating-constraints optimization strategy was found to be only approximately 12%.

We conclude that our method is an important step towards simultaneously achieving the two goals of motion artifact reduction and volume preservation in a fully automated, computationally efficient algorithm.

Acknowledgments

TR was supported by the National Science Foundation under Grant No. EIA-0104114. MAJ was supported by the National Institutes of Health under Grant No. T32 CA09630. All computations were performed on an SGI Origin 3800 supercomputer in the Stanford University Bio-X core facility for Biomedical Computation.

References

1. Tanner, C., Schnabel, J.A., Chung, D., Clarkson, M.J., Rueckert, D., Hill, D.L.G., Hawkes, D.J.: Volume and shape preservation of enhancing lesions when applying non-rigid registration to a time series of contrast enhancing MR breast images. In Delp, S.L., DiGioia, A.M., Jaramaz, B., eds.: Medical Image Computing and Computer Assisted Intervention – MICCAI 2000. Volume 1935 of Lecture Notes in Computer Science., Berlin, Springer-Verlag (2000) 327–337
2. Rohlfing, T., Maurer, Jr., C.R., Bluemke, D.A., Jacobs, M.A.: Volume-preserving non-rigid registration of MR breast images using free-form deformation with an incompressibility constraint. IEEE Trans. Med. Imag. (2003) In press.
3. Rohlfing, T., Maurer, Jr., C.R.: Intensity-based non-rigid registration using adaptive multilevel free-form deformation with an incompressibility constraint. In Niessen, W., Viergever, M.A., eds.: Proceedings of Fourth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2001). Volume 2208 of Lecture Notes in Computer Science., Berlin, Springer-Verlag (2001) 111–119
4. Thirion, J.P.: Image matching as a diffusion process: An analogy with Maxwell's demons. Med. Image. Anal. **2** (1998) 243–260
5. Studholme, C., Hill, D.L.G., Hawkes, D.J.: Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures. Med. Phys. **24** (1997) 25–35
6. Studholme, C., Hill, D.L.G., Hawkes, D.J.: An overlap invariant entropy measure of 3D medical image alignment. Pattern Recognit. **32** (1999) 71–86
7. Rohlfing, T., Maurer, Jr., C.R.: Non-rigid image registration in shared-memory multiprocessor environments with application to brains, breasts, and bees. IEEE Trans. Inform. Technol. Biomed. **7** (2003) 16–25
8. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Non-rigid registration using free-form deformations: Application to breast MR images. IEEE Trans. Med. Imag. **18** (1999) 712–721

Efficient 3D-3D Vascular Registration Based on Multiple Orthogonal 2D Projections

Ho-Ming Chan and Albert C. S. Chung

Dept. of Computer Science, Hong Kong University of Science and Technology, HK
{hmchan,achung}@cs.ust.hk

Abstract. In this paper, we present a new and efficient multi-modal 3D-3D vascular registration algorithm, which transforms the 3D-3D registration problem into a multiple 2D-3D vascular registration problem. Along each orthogonal axis, projected 2D image from a segmented binary 3D floating volume is compared with maximum intensity projection (MIP) image of the reference volume. At the preprocessing stage of the floating image volume, vessels are segmented and represented by a number of spheres with centers located at the skeleton points of the vessels and radii equal to the distance from the skeleton points to their closest boundary. To generate projected images from the binary 3D volume, instead of using the conventional ray-casting technique, the spheres are projected to the three orthogonal projection planes. The discrepancy between the projected image and the reference MIP image is measured by a relatively simple similarity measure, sum of squared differences (SSD). By visual comparison, we found that the performances of our method and the Mutual Information (MI)-based method are visually comparable. Moreover, based on the experimental results, our method for 3D-3D vascular registration is more computationally efficient than the MI-based method.

1 Introduction

Multi-modal image registration is a key for the integration of the complementary image information available in different modalities by aligning the images. Images acquired by different medical imaging modalities provide useful complementary information. For example, time-of-flight magnetic resonance angiographic (TOF-MRA) or 3D rotational angiographic (RA) images provide anatomical information; phase contrast (PC) MRA images provide flow information in the vessels. As such, matching between TOF- and PC-MRA or RA and PC-MRA images can provide flow and structural information in an integrated 3D volume, which is very useful for the physicians.

To correctly align two images, we need a similarity measure to determine how well an image matches against another image through a hypothesized spatial transformation. Most of the similarity measures can be classified into two categories: feature-based or intensity-based. In Section 1.1, we will briefly introduce these similarity measures, and discuss their advantages and disadvantages. In general, the intensity-based method is more accurate. On the other hand, the feature-based method is more computationally efficient [16]. In this paper, we extend the early version of our work

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 301–310, 2003.

in [3] from 2D-3D to 3D-3D registration, and will also compare our proposed method with the Mutual Information (MI)-based registration method [6], [7], which is one of the most commonly used intensity-based methods in the medical image analysis field.

1.1 Previous Works

Over the past few years, many intensity-based similarity measures have been introduced, for example sum of squared differences (SSD), correlation coefficients (CC), correlation ratio (CR) [15] etc. The most popular and widely used one is the information-theoretic similarity measure, namely MI [6], [7]. It does not require segmentation and only makes use of the assumption of statistical dependence between the two images, and has been successfully applied to many multi-modality combinations.

The feature-based registration algorithms extract features such as surfaces, curves [13], [24], [25] or skeletons [11], [12], [14] at the preprocessing steps. They are faster than intensity-based methods but comparatively less accurate [16]. Also, the extraction of skeletons, for example, can be sensitive to noise [17]. Therefore, registration algorithms using skeletons as the only feature for registration can be sensitive to noise as well. On the other hand, the surface or curve extraction and representation can be complicated and would also affect the registration accuracy.

Recently, Chung and et al. [1], [2] used a pair of precisely registered images to build the expected joint intensity or class histogram and employed KLD [4], [5] to measure the discrepancy between observed and expected joint intensity or class histograms. It makes use of the *a priori* knowledge of the expected joint intensity or class histogram to guide the transformation towards the expected outcome.

2 Our 3D-3D Vascular Registration Algorithm

We will provide some background information in the following subsections before we introduce our registration algorithm. For the conventional MI-based registration method, a joint intensity histogram is built at each iteration during the registration process. Within the overlapping region of the floating and reference images, each voxel at the floating image is transformed to the coordinate frame of the reference image and the corresponding intensity bin of the histogram is incremented. Generation of a histogram using the intensity pairs in the overlapping region can be very time-consuming, particularly when the size of image volume is large. A large portion of the registration computational time can be spent on the building of joint intensity histograms. Therefore, we propose to transform the 3D-3D registration problem into a multiple 2D-3D registration problem and compare the projection images perpendicular to the three orthogonal axes until the registration process terminates. We first generate three MIP images of the reference volume along each of the axes. At each iteration, we generate three projected images from the segmented binary floating volume to compare with the corresponding MIPs generated before. An algorithm for generating binary projected images is described in Section 2.2. In order to generate projected images from a binary volume efficiently, some preprocessing of the floating volume is needed.

2.1 Preprocessing of Floating Volume

Before the registration process, an isotropic volume is built so that the new voxel size is equal to the original in-plane voxel size, and the intensities of voxels between slices are estimated by using the trilinear interpolation. Vessels are then segmented by using the global thresholding method and skeletons are extracted [18] with the modified weighted metric $<1,\sqrt{2},\sqrt{3}>$. (Some advanced segmentation method can be used. But, it is not the focus of this paper.) The skeleton points (within a 7x7x7 window) are connected using the depth first search. The main vessel that we are interested is usually the largest connected component containing most skeleton points. By using the extracted skeleton, the original binary volume can be represented and reconstructed by the set of spheres with the coordinates of skeleton points as centers and their distance transform values (their distances to the closest boundary) as radii [22]. Note that a rectangular region of interest containing the main vessels is selected in order to shorten the computational time.

2.2 Generation of Projected Image from a Binary Volume

In order to mimic the MIP images, the whole projected image is randomly initialized to one of the background intensities of the MIP image. The background intensities of the MIP image can be easily obtained from the boundary pixels.

Given the rigid transformation matrix, the center of each sphere of the main vessel is projected to each of the three projection planes under orthogonal projective geometry. The intensities of all pixels with its distance to the projected center less than the corresponding radius are assigned to 255.

A pair of the projected images generated (a) by projection of spheres and (b) by using the ray-casting technique is shown in Fig. 1. The differences of the resulting images are negligible. Our method not only produces similar image compared to the ray-casting technique, but also it is more efficient. A comparison on computational efficiency between the two methods will be presented in Section 3.1.

Fig. 1. Projected images from a binary volume generated by (a) projection of spheres (left) and (b) the ray-casting technique (right).

2.3 Sum of Squared Differences as Similarity Measure

Vessels in both MIP images and projected binary images have high intensity values. Sum of squared differences (SSD), which is a simple similarity measure, can be employed to measure the difference between a 2D MIP image, denoted by U, and the

corresponding binary projected image from the segmented floating volume, denoted by V. The SSD between the two images is given by

$$SSD(U,V) = \sum_{i \in \Phi} (U(i) - V(i))^2 \tag{1}$$

where Φ represents the image domain. When the two images U and V are precisely aligned, the value of SSD is expected to be minimum. We use the sum of the three SSDs for each axis as the similarity measure to guide our registration process.

It is interesting to point out that, from our experience, the similarity measures: MI and KLD also work well in this problem. However, they require extra computations such as building joint intensity or class histograms, normalizing the histograms and calculating marginal distributions (for MI). Moreover, the bin sizes of the joint intensity histograms may affect the registration result.

2.4 Outline of Registration Process

During the registration process, the floating image is the 3D PC-MRA volume while the reference image is the 3D TOF-MRA volume, which is fixed in the 3D space. At each iteration, three binary projected 2D images, along each orthogonal axes, of the segmented PC-MRA volume for the current pose is generated by the method mentioned in Section 2.2. The goal is to find the optimal rigid transformation $\hat{T}$ by minimizing the value of sum of SSDs (see Equation (1)) between the binary projected images and their corresponding MIP images,

$$\hat{T} = \arg\min_{T} (SSD(U_{xy}, V_{xy}(T)) + SSD(U_{xz}, V_{xz}(T)) + SSD(U_{yz}, V_{yz}(T))) \ . \tag{2}$$

Powell's method [10] is used to iteratively search for the minimum value of the sum of SSDs along each parameter (in 3D, the rigid transformation has three translational and three rotational parameters) using Brent's method [10]. The algorithm halts when the percentage change of the sum of SSDs is below a user-specified threshold. We set this threshold to 0.001% in our program. The flow chart of our registration algorithm is shown in Fig. 2.

3 Results and Comparisons

In this section, we show the results of the 3D-3D rigid vascular registration algorithm presented in Section 2 and also compare its accuracy and computational efficiency with the MI-based method. We used three pairs of TOF- and PC-MRA datasets for this experiment to study the performance of the SSD-based method and selected one of the three pairs (Case 1) for comparison with the MI-based method. The specification of the MRA dataset is listed in Table 1.

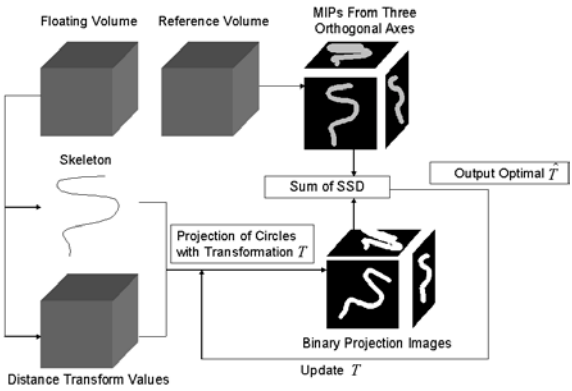


Fig. 2. A flow chart of our registration algorithm

Table 1. Specifications of the datasets used for evaluation of our registration algorithm.

Case No.	Floating Volume		Reference Volume	
	Dimensions (H x W x D)	Voxel Size (mm)	Dimensions (H x W x D)	Voxel Size (mm)
1	512 x 512 x 64	0.390625; 0.390625; 1	512 x 512 x 80	0.390625; 0.390625; 0.75
2	512 x 512 x 32	0.390625; 0.390625; 0.9	512 x 512 x 60	0.410156; 0.410156; 0.95
3	512 x 512 x 64	0.390625; 0.390625; 1	512 x 512 x 60	0.390625; 0.390625; 0.95

3.1 Probing Experiments

The first pair of the volumes was used in this experiment. The resulting transformation obtained by using each method was assumed to be their corresponding optimal transformations. The probing results of the objective functions are shown in Fig. 3. The performances of SSD and MI are comparable when the floating volume was shifted between -10mm to 10mm against Tx, Ty, and was rotated between -18º to 18º.

The computational times of all the evaluations in the probing experiments presented in this subsection are summarized as follows. For the SSD-based method, the mean and standard deviation are 0.19s and 0.02s respectively while for the MI-based method, the mean and standard deviation are 17.79s and 1.41s respectively.

3.2 A Study on Computational Time of Our Registration Algorithm

In order to compare the computational efficiency between our method and the ray-casting method for generating a projected image from binary volume, we generated 1000 projected images from a binary volume using both methods. The ray-casting method takes 1.34 seconds on average whereas our method takes 0.07 seconds on average, which is about 20 times faster than the ray-casting method. The computational times of the whole registration process for SSD-based and the MI-based method are listed in Table 2. All experiments done in this paper were run on a 1.7GHz Pentium IV PC with 768M RAM. Based on the performance of probing experiments and the first dataset, it shows that our method can run faster than the MI-

based method by about 100 times. Note that the MI-based method failed to converge in cases 2 and 3.

Table 2. Comparisons of the computational time (in sec) for the two registration algorithms.

Methods	Case 1	Case 2	Case 3
SSD	36.963	43.192	26.918
MI	4705.746		

3.3 A Study On Performance of Our Registration Algorithm

The results are shown in Fig. 4, which shows the MIPs of the registered MRA volumes using the proposed method. Full MIPs are shown here for the visualization purpose only. During the registration, only binary projected images were generated and used. Segmented vascular regions of the binary projected image (at final alignment using the proposed method) are overlaid on the corresponding DSA images (h) and the results are promising.

4 Discussions and Conclusions

For the preprocessing of 3D volume, the centerline extraction method proposed by Aylward and et al. [26] does not need any segmentation but initial seek points and scales of each vascular branch are required. In our algorithm, instead of using the thinning algorithm [20] or the centerline extraction algorithm [26], skeletons are extracted by using the distance transforms [18], [22]. With the skeleton points and the distance transform values, the original volume can be recovered exactly [22], which may not be feasible by thinning algorithms or centerline extraction algorithms. Skeletons or centerline extraction can be sensitive to noise [17]. Noisy data may give a lot of undesired skeleton points which can affect the accuracy of the method proposed by Aylward and et al. [26] because they use skeletons only. Our method requires skeletonization as well, but it does not use skeletons as the only feature. Instead of using skeletons only, our method also makes use of the distance transform values for generating binary projected images. Thus, the accuracy of our registration algorithm can be less sensitive to noise.

In our projection-based registration method, instead of projecting all vessel voxels, we draw filled circles with centers located at the projected skeleton points, as mentioned in Section 2.2. The reason is that projecting only vessel voxels may result in some blank horizontal or vertical lines in the projected image, which are undesirable and can be avoided by using our method.

To conclude, we have proposed a new and efficient multi-modal 3D-3D vascular registration method based on the efficient generation of projected images from a binary 3D volume. The discrepancy between the projected images and the reference MIP images is measured by the sum of squared differences (SSD) similarity measure. Experimental results show that our method is more computationally efficient than the MI-based method while the accuracies of the both methods are comparable.

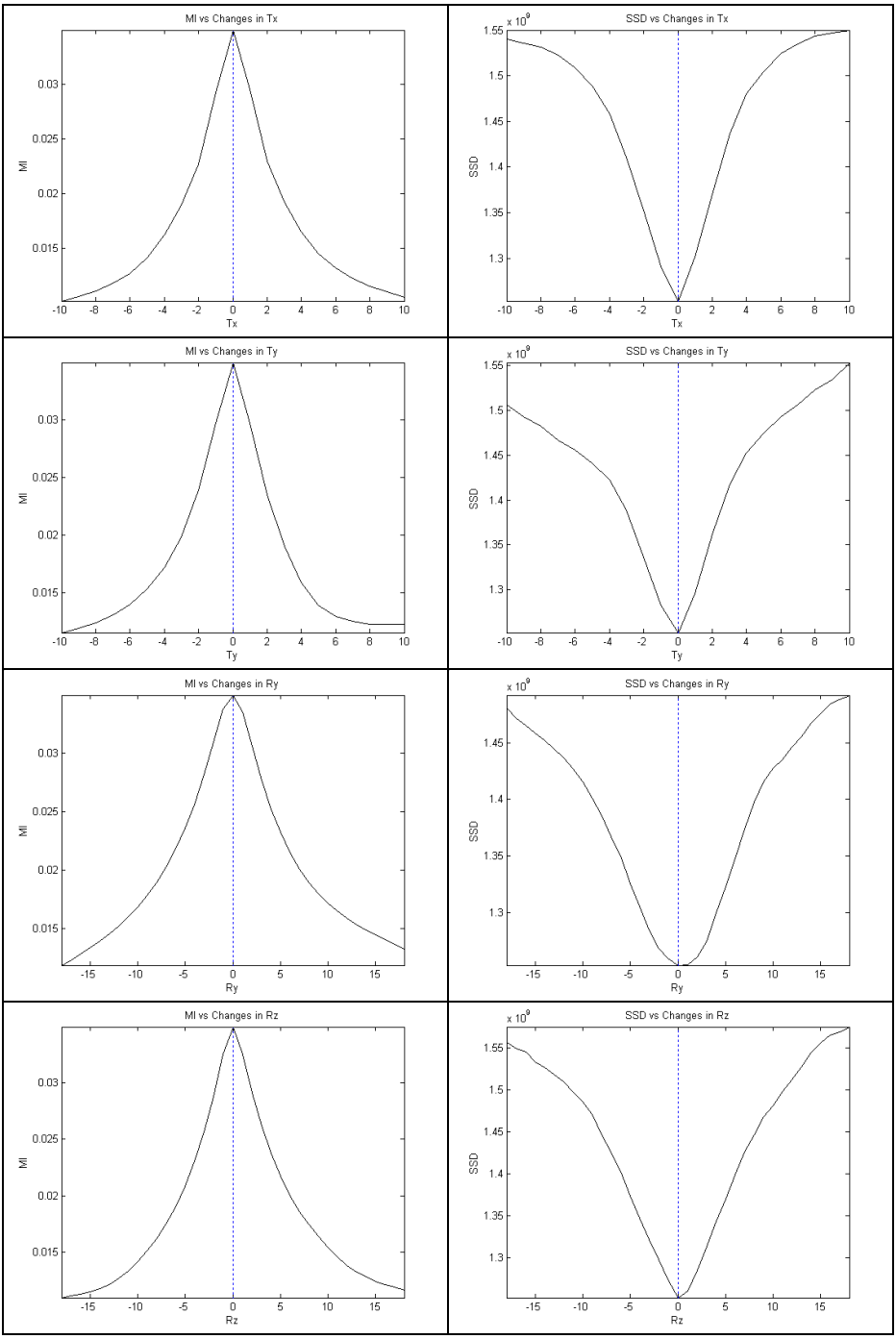


Fig. 3. Probing results of MI(left) and SSD(right) values, in which Tx, Ty, Ry, Rz were varied.

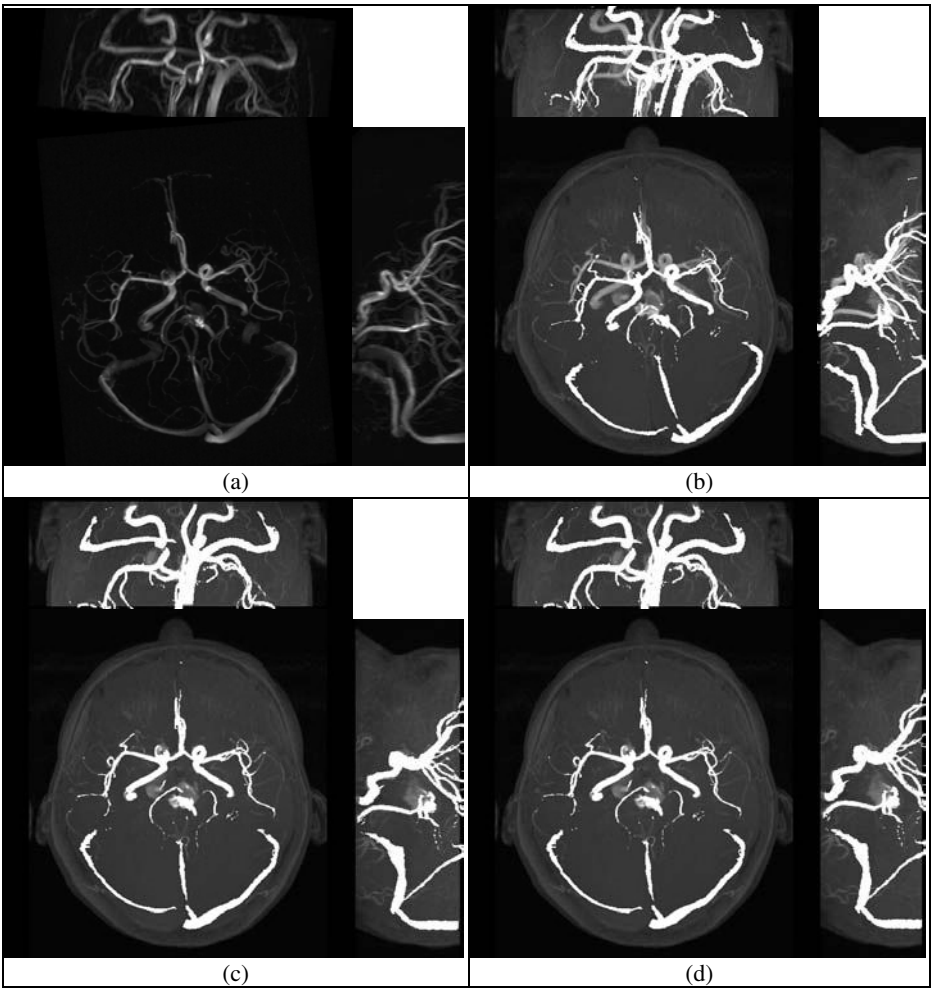


Fig. 4. Results of case 1. (a) MIPs of PC-MRA volume at the initial alignment. (b) Initial image alignment, and binary projected images of PC-MRA volume are overlaid to the TOF-MRA MIP images. (c) Final image alignment by using the SSD-based method, and binary projected images of PC-MRA volume are overlaid to the TOF-MRA MIP images. (d) Final image alignment by using the MI-based method, and binary projected images of PC-MRA volume are overlaid on the TOF-MRA MIP images.

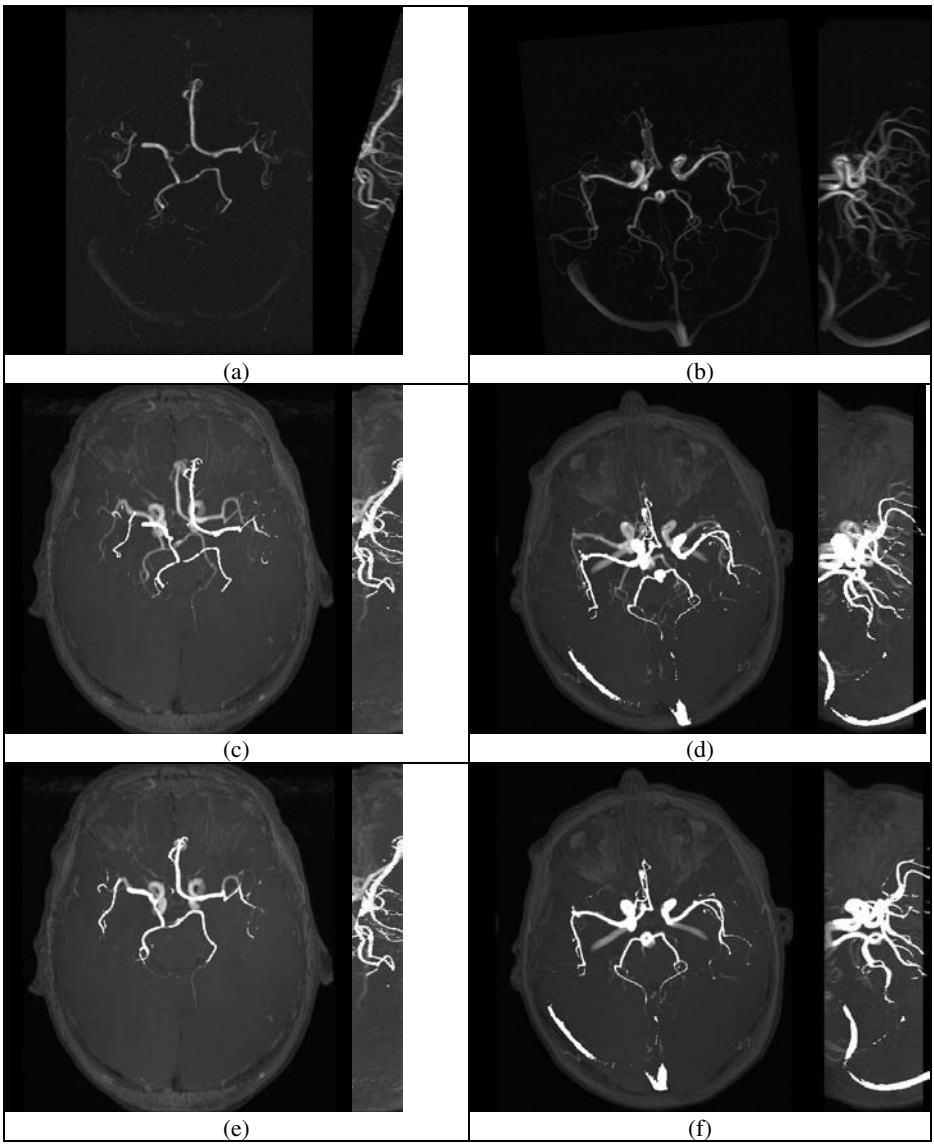


Fig. 5. Results of cases 2 and 3. (a, b) MIPs of PC-MRA volume at initial alignment respectively. (c, d) Initial image alignment, and binary projected images of PC-MRA volume are overlaid to the TOF-MRA MIP images respectively. (e, f) Final image alignment by using the SSD-based method, and binary projected images of PC-MRA volume are overlaid on the TOF-MRA MIP images respectively.

References

1. Albert C.S. Chung, William M. Wells III, and et al., "Multi-modal Image Registration by Minimising Kullback-Leibler Distance", *MICCAI*, 2002, p. 525-532.

2. Ho-Ming Chan, Albert C.S. Chung and et al., “Multi-modal image registration by minimizing Kullback-Leibler distance between expected and observed joint class histograms”, *To appear in CVPR’*03.
3. Ho-Ming Chan, Albert C.S. Chung and Simon C.H. Yu, “2D-3D Vascular Registration Algorithm between Digital Subtraction Angiographic (DSA) and Magnetic Resonance Angiographic (MRA) Images”, *Submitted to MICCAI*, 2003.
4. T.M. Cover and J.A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., ‘91.
5. S. Kullback. *Information Theory and Statistics*. Dover Publications, Inc., 1968.
6. W. Wells, P. Viola, and et al., “Multi-modal Volume Registration by Maximization of Mutual Information”, *Medical Image Analysis*, 1996; 1:32-52.
7. A. Colignon et al., “Automated multi-modality image registration based on information theory”, *IPMI*, p.263-274; Kluwer Academic Publisher, 1995.
8. P. Wyatt and J. Noble, “MAP MRF Joint Segmentation and Registration”, *MICCAI’*02, p.580-587.
9. E. Cosman, “Rigid Registration of MR and Biplanar Fluoroscopy”, *Master Thesis, Dept. of Electrical Engineering and Computer Engineering, MIT,* 2000.
10. W.H. Press, S.A. Teukolsky, and et al., *Numerical Recipes in C*, 2nd Edition, Cambridge University Press, p.402-405,412-420; Kluwer Academic Publisher, 1995.
11. Y. Kita, D.L. Wilson and et al., “Real-time Registration of 3D Cerebral Vessels to X-ray Angiograms”, *MICCAI*, 1997, p.1125-1133.
12. A. Liu, E. Bullitt and et al., “3D/2D Registration via skeletal near projective invariance in tubular objectives”, *MICCAI*, 1998, p.952-963.
13. J. Feldmar, N. Ayache and et al., “3D-2D projective registration of free-form curves and surfaces”, *INRIA*, Report # 2434, 1994.
14. J. Feldmar, G. Malandain and et al., “Matching 3D MR Angiography Data and 2D X-ray Angiograms”, CVRMed-MRCAS'97, Grenoble, France, p.129-138,
15. A. Roche, G. Malandain, and et al., “Multimodal image registration by maximization of the Correlation Ratio”, *INRIA*, Report # 3378, 1998.
16. R.A. McLaughlin, J.Hipwell, and et al., “A comparison of 2D-3D intensity-based registration and feature-based registration for neurointerventions”, *MICCAI*, 2002, p.517-524.
17. R. Fisher, S. Perkins, and et al., http://www.dai.ed.ac.uk/HIPR2/skeleton.htm, 2000.
18. N. Gagvani, “Skeletons and Volume Thinning in Visualization”, *MS. Thesis, Dept. of Electrical and Computer Engineering, Rutgers University, New Brunswick, New Jersey*, 1997.
19. A. Shahrokni, H. Soltanian-Zadeh and et al., “Fast skeletonization algorithm for 3-D elongated objects”, *Proceedings of SPIE*, 2001; 4322:323-330.
20. K. Palagyi, R. Sorantin and et al., “A Sequential 3D Thinning Algorithm and Its Medical Applications”, *IPMI*, 2001, p. 409-415.
21. Y. Zhou, A. Toga, “Efficient Skeletonization of Volumetric Objects”, *IEEE Transactions on Visualization and Computer Graphics*, 1999, p.196-209.
22. A. K. Jain, *Fundamentals of Digital Image Processing*, p.381-389, Prentice Hall, 1989.
23. A. Watt, M. Watt, *Advanced Animation and Rendering Techniques: Theory and Practice*, Addison-Wesley, Reading, Massachusetts, 1992.
24. X. Pennec, “Toward a generic framework for recognition based on uncertain geometric features”, *Videre: Journal of Computer Vision Research*, 1(2):p.58-87, 1998.
25. C. R. Maurer, Jr., R. J. Maciunas, J. M. Fitzpatrick, “`Registration of head CT images to physical space using multiple geometrical features”, *Proc. SPIE Medical Imaging 98*, San Diego, CA , vol. 3338, 72-80 (Feb 1998).
26. Aylward S, Bullitt E, "Initialization, Noise, Singularities, and Scale in Height-Ridge Traversal for Tubular Object Centerline Extraction", *IEEE Transactions on Medical Imaging*, Feb, 2002, Pages 61-75

Semi-automatic Image Registration of MRI to CT Data of the Prostate Using Gold Markers as Fiducials

Jorn A. van Dalen, H.J. Huisman, A. Welmers, and J.O. Barentsz

University Medical Center St. Radboud, Department of Radiology
Geert Grooteplein-Zuid 10, Postbus 9101, 6500 HB Nijmegen, The Netherlands
J.vanDalen@rad.umcn.nl

Abstract. An accurate image registration of MRI to CT data is of great benefit for prostate treatment planning in image-guided radiotherapy. However, prostate motion with respect to surrounding structures and absence of clear structures in and of the prostate in the CT image, complicate this registration. Therefore, we developed a semi-automatic method for a robust, accurate and time efficient image registration of MRI to CT data of the prostate, making use of gold markers, that are inserted in the prostate, as fiducials. The gold markers, as they appear on both images, are segmented using a thresholding technique. The registration deals with a rigid transformation and is based on the iterative closest point algorithm that acts onto the surfaces of the segmented gold markers. The method is clinically tested and uncertainties of 0.4-0.5 mm and 0.7-1.4 mm are obtained at the centre and at the rim of the prostate, respectively.

1 Introduction

Modern medicine frequently employs several imaging techniques within a single patient's case. These different modalities show different, complementary and (or) partially overlapping aspects of the anatomy examined, or show functional aspects, giving few anatomical pointers. Consequently, several medical specialities might benefit from combining images stemming from two or even more modalities.

In many cases, proper integration of different information facilitates correct clinical diagnosis or treatment, see, e.g., [1, 2] . The determination of the geometrical transformation of one of the acquired images to fit another one, i.e., registration, is the first step in this integration process. The second step of the integration is the fusion, required for the integrated display of the data involved. This mainly concerns an adequate visualization.

Since the use of intensity modulated radiation therapy (IMRT), precise external beam radiotherapy treatments can be given to patients [3,4]. In case of the prostate, functional magnetic resonance imaging (MRI), by means of dynamic contrast enhanced MRI (DCE-MRI), using an endo-rectal coil with balloon, H-MR Spectroscopic Imaging (MRSI) and a combination of these two modalities,

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 311–320, 2003.

have been shown to give precise tumor localization [5,6,7,8,9,10,11]. Functional MRI can thus be used for treatment planning in image-guided radiotherapy. However, to do so, functional MRI data have to be integrated with computed tomography (CT) data, since CT-density information is used to calculate the dose distribution during the treatment planning.

The accuracy required from the image registration procedure can be derived from a consideration of the role of geometrical uncertainties in all steps of the radiotherapy process [12, 13], e.g., uncertainties due to prostate position variations and due to patient set-up variations. To account for these uncertainties in the radiotherapy process, a margin is chosen around the clinical target volume. The accuracy of the registration should thus be such that no noticeable margin increase is necessary. Investigation of the patient set-up variations [14] and prostate position variations by means of daily portal imaging of implanted fiducial gold markers [15], led to an accuracy requirement for the image registration of about 2 mm [15]. Furthermore, the geometrical accuracy required in treatment planning and dose delivery of IMRT is also typically 2 mm [16]. Therefore, the accuracy of the registration should be well below this value.

Since prostate motion with respect to surrounding structures is present [17, 18, 19], such an accurate registration of MRI to CT data of the prostate, using these surrounding structures, e.g., bony structures, is not available. Therefore, a registration method of MRI to CT data of the prostate is wanted that does *not* make use of surrounding structures. The main difficulty, however, is the absence of clear structures in the CT image of and in the prostate. In [15], a clinically tested three-dimensional image registration of MRI to CT data of the prostate is presented, making use of gold markers, that were inserted in the prostate, as fiducials. Using these markers for the registration of CT and MRI images in the radiation treatment planning of localized prostate cancer, was also suggested by [20]. The registration performed in [15] is based on manually segmented markers giving accurate results, i.e., uncertainties less than about 2 mm, in most cases. However, in some cases big outliers of about 5 to 15 mm are observed. Furthermore, the manual segmentation is time consuming.

Therefore, we developed a *semi-automatic* three-dimensional image registration of MRI to CT data of the prostate, also making use of gold markers, as fiducials. With the limited user interaction we want to improve the accuracy and robustness of the registration and make it more time efficient.

To perform the registration and test its performance, a database of 20 patients is used. A comparison is made to the results of [15].

2 Materials

In this study, similar materials are being used as in the study of [15]. The database consists of images of 20 consecutive patients with histologically confirmed prostate cancer.

Gold markers that were inserted in the prostate serve to measure and to correct the prostate position during all fractions of the radiotherapy treat-

Table 1. Specifications of the CT and the anatomic MRI scans. For all types the pixel size (PS) in the sagittal-coronal plane, slice thickness (ST) in the axial direction, and field of view (FOV) are given. For the CT scan, the X-ray high voltage peak (V) and the dose rate (R) are given; For the MRI scans the repetition time (TR), echo time (TE) and flip angle (FA) are given. When there is no unique value, a range is given.

modality	PS (mm^2)	ST (mm)	FOV (mm)	V (kVp)	R (mAs)
CT	0.69-0.95	3	354-486	130-140	150-170

modality	PS (mm^2)	ST (mm)	FOV (mm)	TR (ms)	TE (ms)	FA (deg)
MRI T2	0.55	5	280	3500-4400	132	180
MRI T2*	0.56-0.78	3	200-285	699-1100	18-25.8	30

ment [19, 21]. Gold markers were chosen, since they can be visualized in the portal images made during radiotherapy treatment using the transmitted high-energy photon beam of the linear accelerator. About two weeks before performing the imaging studies, on average four gold markers, but not more than five or less than three, with a length of 5 or 10 mm and a thickness of 1 mm, were inserted through trans-rectal ultrasound guidance by an urologist. Usually, two gold markers were inserted at the base, one at the apex and one at the centre of the prostate. In this study, the markers were used as fiducials for the image registration.

The imaging protocol consisted of two examinations. Firstly, CT was performed using a multi-slice spiral CT scanner (Marconi AcQsim, Cleveland, USA) with specifications as given in Table 1. In twelve patients an endo-rectal coil with balloon, with a similar design as the commercially available MRI endo-rectal coil (Medrad, Pittsburgh, US), was inserted and inflated with 80 cc of air, before the CT scan was taken.

Secondly, MRI was conducted using a 1.5 T scanner (Siemens Sonata, Erlangen, Germany). T2-weighted turbo spin echo (TSE) sequences in three planes were acquired. A T2*-weighted sequence (a multi-echo data image combination (MEDIC)) was acquired at the same slice location as the axial T2-weighted TSE sequence. Specifications of the these anatomical MRI scans can be found in Table 1. Apart from the anatomical MRI scans, functional MRI scans were taken. Magnetic resonance spectroscopic imaging was performed using the 3D-PRESS sequence. A multi-slice T1-weighted gadolinium-enhanced FLASH sequence was also performed. For all patients the endo-rectal coil with balloon was inserted before the MRI scans were taken. Using this endo-rectal coil, the MRI image quality improved due to a significant increase of the signal-to-noise ratio. The total examination time took about 60 minutes. Since no significant motion artefacts were observed in this time, it is assumed that accurate matches can be obtained between the functional and anatomic MRI images.

Imaging data were transferred to a PC for post-processing. Using image viewing and registration software, based on Tcl/Tk [22] and the visualization toolkit VTK [23], the gold markers, as they appear on the CT scan as hyper dense and on the T2*-weighted MRI images as low signal intensity areas, were used for

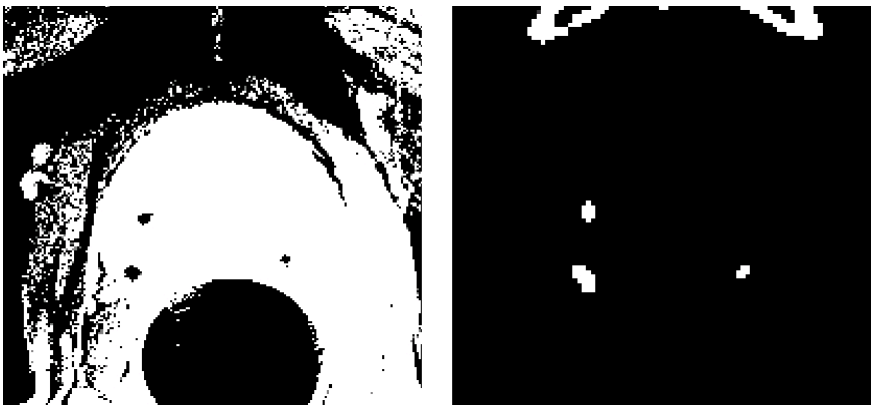


Fig. 1. An MRI and a CT slice of the prostate after thresholding, where three markers are clearly visible.

semi-automatic segmentation and subsequently registration. Due to the length of the markers, their orientation and the slice thickness of both images, the markers are visible in at least one slice, but in most cases in two or three slices.

3 Methods

To achieve a successful registration, first a segmentation of the prostate markers is performed. The semi-automatic segmented markers are used for the registration, where the ICP algorithm is applied onto the surfaces of those segmented markers. The registration is based on a rigid transformation, i.e., based on three translation and three rotation parameters.

3.1 Segmentation

Firstly, the user indicates the centre of the markers, denoted by seeds, $\boldsymbol{s}_m$, where m represents the seed number, as good as possible, on both the MRI and CT images. Since the gold markers are easy to find on the CT image, but less easy on the MRI image, the user can manually align the MRI and the CT image. This will assist him to identify the markers on the MRI image. Both on the MRI as on the CT image, an equal number of seeds is indicated.

Secondly, a binary thresholding is performed on both images. The markers can then be identified by isolated clusters of voxels, with dimensions of about $1.0 \times 0.3 \times 0.3$ cm and with grey values below a threshold T_{MRI} for the MRI image and above a threshold T_{CT} for the CT image. Figure 1 shows an example of an MRI and a CT slice after thresholding, where three markers are clearly visible.

Next, only information is considered within those ellipsoids with centres given by $\boldsymbol{s}_m$ and radii of 1.5 cm in the axial (z) direction and 0.5 cm in both the sagittal

(x) and coronal (y) direction. The set of volumes given by the voxels within the ellipsoids is denoted by V. This mask avoids voxels that are relatively far from the markers, but having about the same grey value as those that compose the marker, to be used in the segmentation.

To obtain the segmented markers, the voxels within the same grey value range (below T_{MRI} for the MRI and above T_{CT} for the CT image) and those that are connected (via neighbouring voxels within the same grey value range) to the user supplied seeds, $\boldsymbol{s}_m$, are marked. Clusters, C_m, of these marked voxels, $v_{m,n}$ (with n the voxel number), then represent the markers, i.e.,

$$C_m = \{v_{m,n} | v_{m,n} \in V\} \quad . \tag{1}$$

Next, triangulated iso surfaces are determined using the boundary value of the clusters as the contour value. Finally, normals are computed for these surfaces, for each of the polygonal facets. The normals are computed via the cross product $[v_1 - v_2] \times [v_2 - v_3]$, where v_1, v_2 and v_3 are three vertices of the polygon, and are averaged at shared points. Using these normals, a faceted shading of the surface is obtained, that will be used as an input of the ICP algorithm. The iso surfaces and the normals are computed using the VTK methods vtkContourFilter and vtkPolyDataNormals [23], respectively.

3.2 Registration

To achieve a successful registration, it is important that first an initial estimate of the relative pose of the two prostate images is given. This is done by roughly aligning the two images, in the region of the prostate, manually. Then, the ICP algorithm performs the final registration step, automatically.

The iterative closest point algorithm [24, 25, 26] is a method for the registration of three-dimensional shapes. It works in terms of registration of collected data, which are converted to a point set $P = \{\boldsymbol{p}_i\}$, of image 1, to model shape data X, which remain in their original representation, of image 2. In this paper, P is a point set located onto the surface of the segmented MRI markers and X is the surface of the segmented CT markers.

The algorithm has three stages and iterates. The first step is finding the closest model point $\boldsymbol{x}_i \in X$ with respect to $\boldsymbol{p}_i$ for all i. The point $\boldsymbol{x}_i \in X$ for which the distance $d(\boldsymbol{p}_i, X)$ between $\boldsymbol{p}_i$ and $\boldsymbol{x}_i$ is minimum, can be derived from

$$d(\boldsymbol{p}_i, X) = \min_{\boldsymbol{x}_i \in X} |\boldsymbol{x}_i - \boldsymbol{p}_i| \quad . \tag{2}$$

For a triangulated surface - the model representation of the image data that is being used - the model X comprises a set of triangles $T = \{t_k\}$. If triangle t_k has vertices $\boldsymbol{r}_{k,j}$ $(j = 1, 2, 3)$, then the minimum distance between $\boldsymbol{p}_i$ and t_k is

$$d(\boldsymbol{p}_i, t_k) = \min_{u+v+w=1} |u\boldsymbol{r}_{k,1} + v\boldsymbol{r}_{k,2} + w\boldsymbol{r}_{k,3} - \boldsymbol{p}_i| \quad , \tag{3}$$

where u, v and w are in $[0, 1]$. Thus, the closest distance between the point $\boldsymbol{p}_i$ and the triangle set T is given by

$$d(\boldsymbol{p}_i, T) = \min_k d(\boldsymbol{p}_i, t_k) \quad . \tag{4}$$

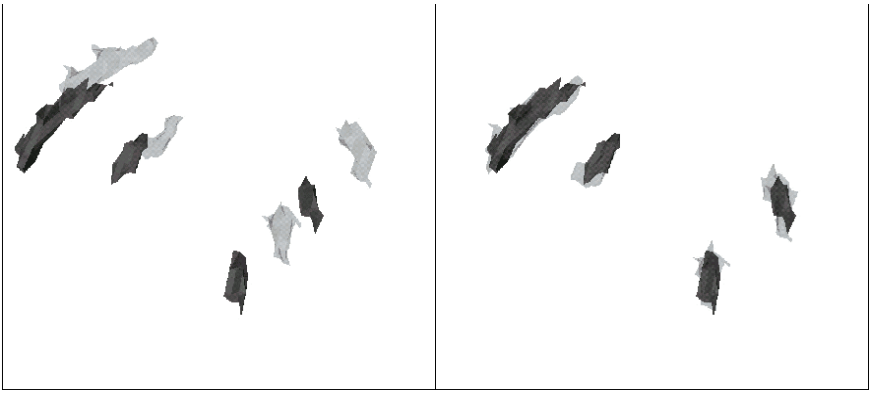


Fig. 2. Four segmented markers in both MRI (dark) and CT (light) images, in 3D, before (left) and after (right) applying the ICP method.

The closest model point to the data point $\boldsymbol{p}_i$ is then given by $\boldsymbol{x}_i = (u\boldsymbol{r}_{k,1}, v\boldsymbol{r}_{k,2}, w\boldsymbol{r}_{k,3})$, where $\boldsymbol{r}_{k,j}$ are the vertices of t_k that satisfies Equation (4).

Secondly, a least square registration between the points $\{\boldsymbol{p}_i\}$ and $\{\boldsymbol{x}_i\}$ can be carried out by minimizing

$$f(\boldsymbol{q}) = \frac{1}{N_\mathrm{p}} \sum_{i=1}^{N_\mathrm{p}} |\boldsymbol{x}_i - R(\boldsymbol{q}_\mathrm{R})\boldsymbol{p}_i - \boldsymbol{q}_\mathrm{T}|^2 \quad . \tag{5}$$

where N_p is the number of point pairs being used, $R(\boldsymbol{q}_\mathrm{R})$ is a rotation matrix and $\boldsymbol{q} = (\boldsymbol{q}_\mathrm{R}, \boldsymbol{q}_\mathrm{T})$, resulting from $\arg\min_{\boldsymbol{q}} f(\boldsymbol{q})$, is the transformation vector consisting of three rotation parameters $\boldsymbol{q}_\mathrm{R} = (\alpha_1, \alpha_2, \alpha_3)$ and three translation parameters $\boldsymbol{q}_\mathrm{T} = (T_x, T_y, T_z)$. The function $f(\boldsymbol{q})$ is a measure of the accuracy of the method: the value of $\sqrt{f(\boldsymbol{q})}$ could be interpreted as a standard deviation, giving a measure of the uncertainty of the matching.

Thirdly, the set of data points $\{\boldsymbol{p}_i\}$ is then transformed to $\{\boldsymbol{p}_i'\}$ using the calculated rigid body transformation. Then, the iteration procedure starts by redetermining the closest point set using $\{\boldsymbol{p}_i'\}$ and X. The algorithm terminates when the change in $f(\boldsymbol{q})$ between two successive iterations falls below a threshold. In this analysis this change is set to 0.0001 mm. It also terminates when the number of iterations exceeds 3000. Furthermore, N_p is set to 5000.

An example of the performance of the ICP method is shown in Figure 2: four segmented markers in both MRI and CT images, in 3D, before and after applying the ICP method are visualized. Typically, the ICP registration takes a few seconds.

3.3 Evaluation

For each patient, the computed ICP registration transformation, M, consisting of three rotation and three translation parameters, is applied onto two points

in the MRI image. The first point, $\tilde{\boldsymbol{c}}$, is the average value of the centre-of-mass points of the segmented MRI markers. This point, which is determined by one user, estimates the centre of the prostate. The second point, $\tilde{\boldsymbol{r}}$, is defined as being 2 cm back and down and to the left, with respect to $\tilde{\boldsymbol{c}}$, which would be in the rectum wall region. The resulting points $\boldsymbol{c} = (c_x, c_y, c_z) = M\tilde{\boldsymbol{c}}$ and $\boldsymbol{r} = (r_x, r_y, r_z) = M\tilde{\boldsymbol{r}}$ are thus estimations of the centre and the rim of the prostate after applying the ICP registration.

Inter- and intra-operator variability of the parameter values c_i and r_i ($i = x, y, z$) were determined, in order to get a measure of the registration uncertainty around the centre and the rim of the prostate. This measure assumes that a rigid transformation is sufficient to register the segmented MRI markers onto the segmented CT markers. Variability can be expected on the basis of the user input: how many and which markers does the user select and what threshold values does he use for the segmentation. One operator performed the segmentation and registration five times for five data sets in order to obtain the intra-operator variability. Variability in the registration between three different operators (the inter-operator variability) was measured via a multivariate analysis of variance (MANOVA) test [29]. Furthermore, the effect of using an endo-rectal balloon in the CT images was investigated. All variabilities are obtained via R [30], a freely available language and environment for statistical computing and graphics.

4 Results

The threshold values that are used in the segmentation are set to $T_{\mathrm{MRI}} = 150$ and $T_{\mathrm{CT}} = 900$, with a possible change by the user of ± 50 and ± 400, respectively. Experimentally, these are found to be good values to isolate the gold markers and thus to perform a marker segmentation in both the MRI and CT images.

Studying the variability of the registration results, the MANOVA test suggests that contributions to the variability of intra-operator and balloon effects are not statistically significant. There is a slight effect ($p < 0.08$) due to repeated registrations by different operators. The contribution of inter-operator variability, however, is small.

Figure 3 shows box plots of the total variability of the coordinates c_i and r_i ($i = x, y, z$) after normalization with respect to the mean, per patient.

The three patients in which the biggest variations are observed are patients 3, 5 and 14. The variations are caused by having missed one marker by one user on the MRI image (patients 3 and 14) and by difficulties in getting a good segmentation of hardly visible MRI markers (patient 5).

The estimated total standard deviations at the centre and at the rim of the prostate are given in Table 2. The uncertainties of [15] are also shown.

Our results indicate that the uncertainties are well below the required limit of 2 mm. Furthermore, they are slightly better than the results, excluding the outliers, of [15]. When the outliers observed in [15] are not excluded, our results are much more accurate.

The limited user interaction in our method prevents big outliers to occur and thus makes the method robust and accurate. Moreover, it results in a time effi-

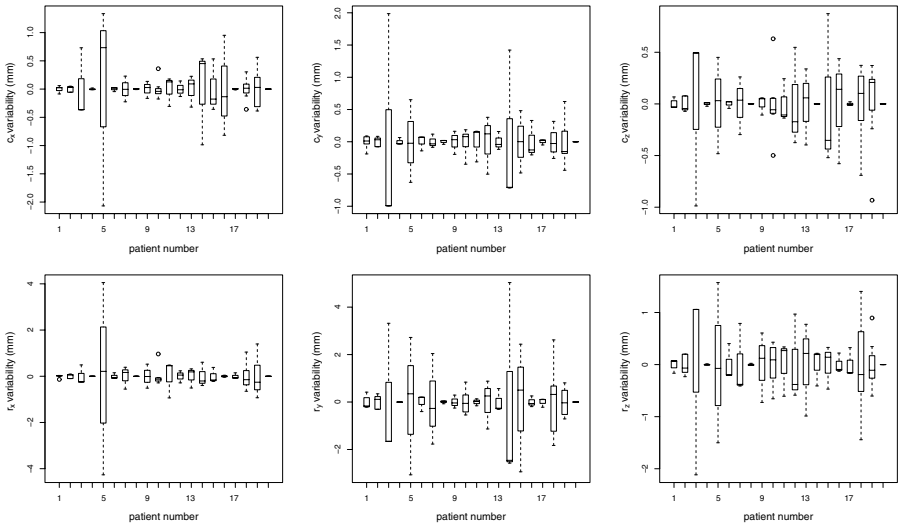


Fig. 3. Box plots of the variability of the coordinates c_i and r_i $(i = x, y, z)$ after normalization with respect to the mean, per patient.

Table 2. Uncertainties, in mm, at the centre, (c_x, c_y, c_z), and at the rim, (r_x, r_y, r_z), of the prostate. Our results and those of [15] are given.

reference	c_x	c_y	c_z	r_x	r_y	r_z
our results	0.5	0.5	0.4	0.9	1.4	0.7
[15] without outliers	0.4	0.6	0.5	1.3	1.5	0.8
[15] with outliers	1.5	2.9	0.9	1.7	3.0	1.1

cient registration: on average, the registration time, from the moment of loading the images till the moment of saving the registration, took about 8 minutes, a factor three less than the time given in [15]. Most likely, the gain in time is due the automation of the segmentation.

5 Conclusions and Discussions

In this paper a semi-automatic method for image registration of MRI to CT data of the prostate is presented, making use of gold markers, that are inserted in the prostate, as fiducials.

The method is clinically tested and uncertainties of 0.4-0.5 mm and 0.7-1.4 mm are obtained at the centre and at the rim of the prostate, respectively. These results are well within the clinical required accuracy for treatment planning in image-guided radiotherapy.

The limited user interaction results in a time efficient registration and prevents big outliers to occur and therefore makes the method robust and accurate. In particular, this is true when comparing to [15], where a similar analysis is

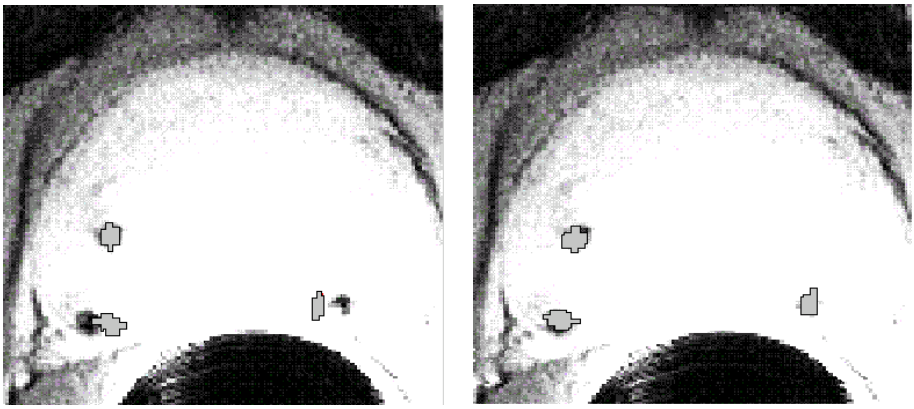


Fig. 4. An extreme example of an MRI and a CT slice of the prostate after ICP registration, where a rigid transformation clearly does not suffice (left), but where an affine transformation gives a much better result (right). The window level of the CT image is such that only the gold markers (indicated by three grey bounded areas) are visible.

performed, but, where the segmentation of the gold markers is performed manually.

In this paper, the assumption is made that a rigid transformation can register the segmented MRI markers onto the segmented CT markers. Thus, it is assumed that imaging artefacts, differences in deformations of the prostate and changes in distance between the markers on CT and MRI images due to marker migration, are negligible and consequently do not affect the registration. This is a strong assumption and found not to be valid for at least some of the patient's images. This is illustrated by the extreme example of Figure 4 where an MRI and CT slice of an image volume of the prostate are shown after applying the ICP registration based on a rigid transformation. A rigid transformation clearly does not suffice here. In 4 (out of 20) patients, we visually observe that a rigid transformation is not able to match the MRI and CT markers. Using an affine transformation, however, where we add three scaling parameters in the transformation, i.e., where we replace $\boldsymbol{q}_{\mathrm{R}}$ of Equation (5) by a vector consisting of 6 elements (three for rotation and three for scaling), we visually observe a much better result. This is also shown in Figure 4.

It has to be investigated further what the registration uncertainties are, quantitatively, as a consequence of assuming that a rigid transformation can register the segmented MRI markers onto the segmented CT markers. Next, it needs to be investigated how non-rigid transformations can improve the registration.

References

1. J.B.A. Maintz and M.A. Viergever, Medical Image Analysis **2** nr. 1 (1998), 1.
2. D.L.G. Hill et al., Phys. Med. Biol. **46** (2001), 1.

3. C. Burman et al., Int. J. Radiat. Oncol. Biol. Phys. **39** nr. 4 (1997), 863.
4. M.J. Zelefsky et al., Radiother. Oncol. **55** nr. 3 (2000), 241.
5. M. Engelbrecht et al., *Dynamic MR imaging of prostate cancer: dynamic features for cancer detection* (2002), submitted to Radiology.
6. M. van der Graaf et al., MAGMA **10** nr. 1 (2000), 153.
7. Y. Kaji et al., Radiology **206** nr. 3 (1998), 785.
8. J. Scheidler, Radiology **213** nr. 2 (1999), 473.
9. J. Kurhanewicz, Radiol. Clin. North. Am. **38** nr. 1 (2000), 115.
10. U. Mueller-Lisse et al., Radiology **221** nr. 2 (2001), 380.
11. F. van Dorsten et al., Radiology **221** nr. P (2001), 585.
12. J. C. Stroom et al., Int J. Radiat. Oncol. Biol. Phys. **43** nr. 4 (1999), 905.
13. M. van Herk et al., Int J. Radiat. Oncol. Biol. Phys. **52** nr. 5 (2002), 1407.
14. E.N. van Lin et al., Int. J. Radiat. Oncol. Biol. Phys. **50** nr. 2 (2001), 569.
15. A. Welmers et al., *3D Image Fusion of CT with Dynamic Contrast Enhanced MRI and MR Spectroscopy of the Prostate Using Gold Markers as Fiducials*, Radiological Society of North America (RSNA) Abstracts, Suppl. to Radiology **225** nr. P (2002), 350. Presented at RSNA 2002.
16. M. Partridge et al., Phys. Med. Biol. **45** nr. 12 (2000), N183.
17. M. van Herk et al., Int. J. Radiat. Oncol. Biol. Phys. **33** nr. 5 (1995), 1311.
18. M.J. Zelefsky et al., Radiother. Oncol. **50** nr. 2 (1999), 225.
19. H. Alasti et al., Int. J. Radiat. Oncol. Biol. Phys. **49** nr. 3 (2001), 869.
20. C.C. Parker et al., Radiother. Oncol. **66** nr. 2 (2003), 217.
21. J.M. Balter, Int. J. Radiot. Oncol. Biol. Phys. **31** nr. 1 (1995), 113.
22. `http://www.tcl.tk/`
23. The Visualization Toolkit User's Guide, ISBN 1-930934-08-4, published by Kitware; The Visualization Toolkit, An Object-Oriented Approach To 3D Graphics, 3rd edition, ISBN 1-930934-07-6, published by Kitware; `http://public.kitware.com/VTK/`
24. P.J. Besl and N.D. McKay, IEEE Transactions on Pattern Analysis and Machine Intelligence **14** nr. 2 (1992), 239.
25. C. Chen and G. Medioni, Image Vision Computing **10** nr. 3 (1992), 145.
26. Z. Zhang, International Journal of Computer Vision **13** nr. 2 (1994), 119.
27. D. Litzenberg et al., Int J. Radiat. Oncol. Biol. Phys. **52** nr. 3 (2002), 699.
28. H. Alasti et al., Int J. Radiat. Oncol. Biol. Phys. **49** nr. 3 (2001), 869.
29. see, e.g., D.F. Morrison, *Multivariate Statistical Methods*, McGraw-Hill (1967), New York; `http://online.sfsu.edu/~efc/classes/biol710/manova/manovanew.htm`
30. `http://www.R-project.org/`

Image Registration for Interventional MRI Guided Procedures: Interpolation Methods, Similarity Measurements, and Applications to the Prostate

Baowei Fei[1], Zhenghong Lee[1,2], Jeffery L. Duerk[1,2], and David L. Wilson[1,2]

[1]Department of Biomedical Engineering
Case Western Reserve University, Cleveland, OH 44106, USA
{BXF18,DLW}@po.cwru.edu
[2]Department of Radiology, University Hospitals of Cleveland, Cleveland, OH 44106, USA
{Lee,Duerk}@uhrad.com

Abstract. Nuclear medicine can detect and localize tumor in the prostate not reliably seen in MR. We are investigating methods to combine the advantages of SPECT with interventional MRI (iMRI) guided radiofrequency thermal ablation of the prostate. Our approach is to first register the low-resolution functional images with a high resolution MR volume. Then, by combining the high-resolution MR image with live-time iMRI acquisitions, we can, in turn, include the functional data and high-resolution anatomic information into the iMRI system for improved tumor targeting. In this study, we investigated registration methods for combining noisy, thick iMRI image slices with high-resolution MR volumes. We compared three similarity measures, i.e., normalized mutual information, mutual information, and correlation coefficient; and three interpolation methods, i.e., re-normalized sinc, tri-linear, and nearest neighbor. Registration experiments showed that transverse slice images covering the prostate work best with a registration error of ≈ 0.5 mm as compared to our volume-to-volume registration that was previously shown to be quite accurate for these image pairs.

1 Introduction

Nuclear medicine can detect and localize tumor in the prostate not reliably seen in MR [1]. We are investigating methods to combine the advantages of SPECT with interventional MRI (iMRI) guided radiofrequency (RF) thermal ablation for the treatment of the prostate cancer. Our idea is to first register the low-resolution functional images with a high resolution MRI [2]. Then by registering the high-resolution MR volume with live-time iMRI acquisitions, we can, in turn, map the functional data and high-resolution anatomic information to iMRI images to aid tumor targeting. We previously reported a method for the registration of noisy, thick iMRI image slices with high-resolution MR volumes with simulated [3] and actual iMRI images [4]. In this report, we compared three interpolation methods and three similarity measures for this application. Hundreds of registration experiments were performed with 12 pairs of MR volume images acquired from four healthy volunteers.

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 321–329, 2003.

2 Registration Algorithms

2.1 Three Interpolation Methods

We investigated three interpolation methods, i.e., re-normalized sinc interpolation [5], tri-linear, and nearest neighbor.

Let the original data set be I_{org}, the re-formatting data set I_{new}. The conventional sinc interpolation with a cosine Hamming window is described as below [5,6].

$$I_{new}(x,y,z)=\sum_{X}\sum_{Y}\sum_{Z} I_{org}(X,Y,Z)\cdot H(x,X,R)\cdot H(y,Y,R)\cdot H(z,Z,R)$$

$$\text{where } H(a,A,R)=\frac{\sin[\pi(a-A)]}{2\pi(a-A)}\cdot\left\{1+\cos[\pi(a-A)/R+1]\right\}.$$

and *X, Y, Z*, represent the coordinates of (I_{org}); *x, y, z* the coordinates of (I_{new}); *A* is a symbol representing *X, Y*, or *Z,* and *a* represents *x, y*, or *z;* and *R* is the kernel size. The Hamming function eliminates problems with oscillatory effects at discontinuities and guarantees that the convolution coefficients fall of to zero at the edge of the sinc kernel (i.e., at $|a|=R+1$, where $R=5$ in this study) [5].

In our implementation, we used the re-normalized sinc interpolation method because it could make significant improvement in performance of the conventional sinc interpolation [5]. We replaced *H* in the above equation with

$$H_{new}(a,A,R)=H(a,A,R)\Big/\sum_{A}H(a,A,R)=H(a,A,R)\Big/\left|H(a)\right|.$$

2.2 Three Similarity Measurements

We used three similarity measures, normalized mutual information (NMI), mutual information (MI), and correlation coefficient (CC), in our registration. One image *R* is the *reference*, and the other *F* is *floating*. Their mutual information *MI* is given below [7,8].

$$MI=\sum_{r,f} p_{RF}(r,f)\log\frac{p_{RF}(r,f)}{p_R(r)\cdot p_F(f)}$$

The joint probability $p_{RF}(r,f)$ and the marginal probabilities $p_R(r)$ of the reference image and $p_F(f)$ of the floating image, can be estimated from the normalized joint intensity histograms. We used the NMI version proposed by Maes [8].

$$NMI=\frac{2MI}{H(R)+H(F)}$$

where $H(R)=-\sum_{r} p_R(r)\log p_R(r)$ and $H(F)=-\sum_{f} p_F(f)\log p_F(f)$.

The correlation coefficient CC is given below [9].

$$CC = \frac{\sum (R(r) - \overline{R})(F(f) - \overline{F})}{\sqrt{\sum (R(r) - \overline{R})^2 \sum (F(f) - \overline{F})^2}}$$

Here $\overline{R}$, $\overline{F}$ denote the average intensities of the reference and floating images and the summation includes all voxels within the overlap of both images.

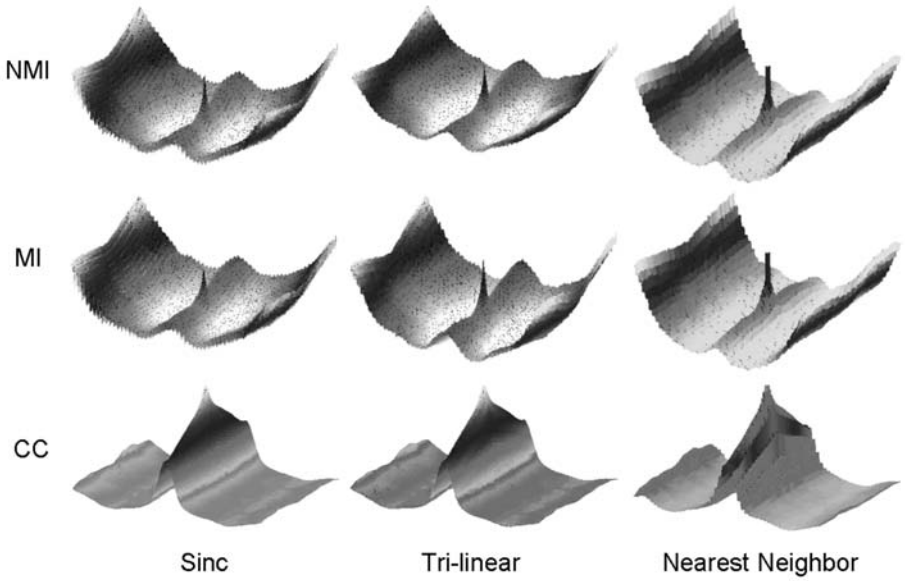


Fig. 1. Similarity surfaces are plotted as a function of translations at the 1/4 resolution in the multi-resolution registration process. Two high-resolution MRI volumes were registered, and they are down sampled by 1/4 along each linear dimension, giving a distance between voxel centers of ≈ 5.5 mm. From the optimal parameters, we computed the similarity values of the simulated iMRI and MRI images as a function of translations along the coronal (anterior-posterior) and sagittal (left-right) axis. From top to bottom, normalized mutual information (NMI), mutual information (MI), and correlation coefficient (CC) surfaces are plotted. From left to right, sinc, tri-linear, and nearest neighbor interpolations are used to obtain the floating images, respectively. The noisey NMI/MI surfaces show a false global maximum and many local maxima. CC surfaces are much smoother indicating its suitability for low resolution. Nearest neighbor has a flat peak with a width of one voxel in similarity surfaces. Images are from volunteer S2.

2.3 Comparison of Similarity Surfaces

We plot the similarity surfaces for the three similarity measures, NMI, MI, and CC, at different resolutions; and we determine their suitability for SV registration. At *1/4 resolution*, we resampled images so as to give 1/4 number of the voxels along each linear dimension. At *full resolution*, we used the full number of voxels. We plot the similarity measures as a function of translations. After two typical high-resolution

MR volumes were registered [10], values were plotted with the origin as the optimal transformation. We calculated similarity values while moving the simulated iMRI image relative to the high-resolution MR image along coronal (anterior-posterior) and sagittal (left-right) axis. When obtaining floating images, we used the three different interpolation methods.

At 1/4 resolution (Fig. 1), CC surfaces are much smoother than NMI and MI, which are noisy and contain a false global maximum that could lead to a false answer and many local maxima [11]. From these figures, we infer that CC is better at low resolution. Comparing CC surfaces of different interpolations, sinc and tri-linear have similar surfaces, and tri-linear is better than nearest neighbor. For this application, we chose tri-linear interpolation instead of sinc because it is much faster and because it has comparable performance. Finally, we used CC and tri-linear at low resolution.

At full resolution (Fig. 2), NMI and MI surfaces are much more peaked than CC that infers good optimization accuracy, but once again there is high frequency noise in the NMI and MI curves, far from the optimum, that gives rise to local maxima that must be avoided. Comparing three interpolation methods, sinc gave the sharpest peak at the optimum; nearest neighbor interpolation gave a flat peak with a width of one voxel; and tri-linear gave a result between the other two. As stated above, tri-linear is much faster that sinc with similar performance. NMI and MI have no significant difference but NMI is a little bit robust in our implementation. We chose NMI and tri-linear at the full resolution.

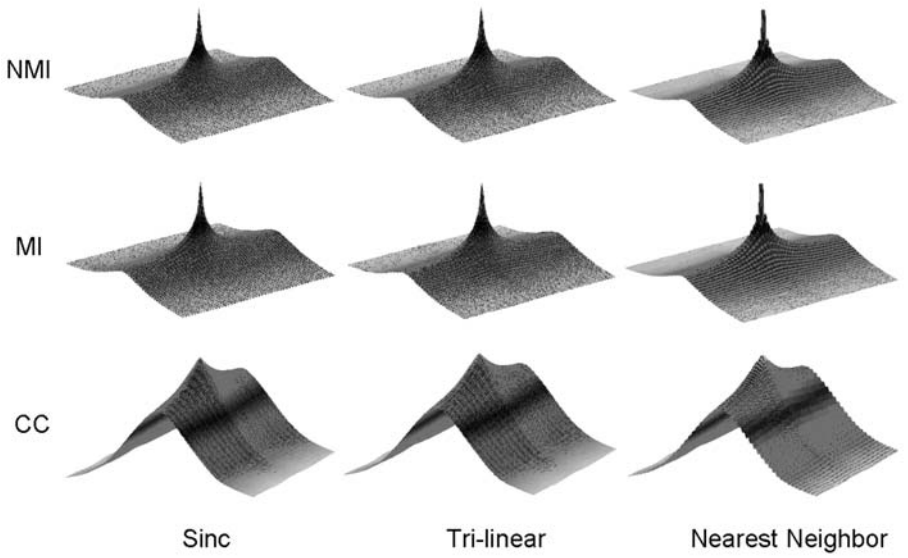


Fig. 2. Similarity functions are plotted as a function of translations at full resolution. Many details are given in the legend of Fig. 1. NMI and MI surfaces are much peaked than CC, especially with sinc and tri-linear interpolation. The voxel is isotropic with 1.4 mm on a side. Image data are the same used in Fig. 1.

2.4 Combination of Normalized Mutual Information and Correlation Coefficient

As a result of the above analyses, we created a registration algorithm for prostate MR images. We define the iMRI image slice to be the *reference* image*;* the matching slice from the high-resolution MRI volume is the *floating* image. We use a multi-resolution approach and perform registration from low to high resolution. We use CC at the two lower resolutions because it gives fewer local maxima and because it can be calculated faster than NMI. We use NMI at full resolution because of its peaked surface. To avoid local maxima, we include a restarting feature where registration is restarted with randomly perturbed parameters obtained from a uniform distribution about the initial transformation values at the current resolution being used. The algorithm restarts until the absolute CC is above a threshold of 0.5 as experimentally determined or the maximum number of restarts is reached.

For registration, we use a rigid body transformation (three translations and three rotations). For optimization, we use the downhill simplex method of Nelder and Mead [12]. Optimization of similarity ends either when the maximum number of calculations is reached (typically 500) or when the fractional change in the similarity function is smaller than a tolerance (typically 0.001). We use IDL (Interactive Data Language, Research System Inc., Boulder, CO.) as the programming language. We use an initial guess assuming an identity transformation, i.e., all initial translation and rotation parameters are zero, because the patient is normally oriented approximately the same way from one scan to the next. We set the maximum numbers of restarts at 10, 5, and 3, from low to high resolution, respectively.

2.5 Registration Evaluation

We used a variety of evaluation methods. We used *RegViz,* a program created in IDL in our laboratory with multiple visualization and analysis methods. First, we manually segmented prostate boundaries in image slices and copied them to corresponding slices. This enabled visual determination of the overlap of prostate boundaries over the entire volume. Second, color overlay displays were used to evaluate overlap of structures. To visualize potential differences, it was quite useful to interactively change the contribution of each image using the transparency scale. Third, we used a sector display, which divided the reference and registered images into rectangular sectors and created an output image by alternating sectors from the two input images. Even subtle shifts of edges would be clearly seen.

Our quantitative evaluation method for slice to volume registration was to compare SV and VV registration [10]. For volume pairs acquired over a short time span from a supine subject with legs flat on the table, prostates were well aligned and prostate centroid displacements were typically < 1 mm. The registration accuracy as determined from displacements of pelvic bony landmarks was 1.6 ± 0.2 mm, a value comparable to error associated with locating the landmarks [10]. To compare SV and VV registration, we defined a rectangular volume of interest (VOI) just covering the prostate and calculated voxel displacements between the two registrations. We defined the SV registration as being *successful* when the 3D displacement was less than 2.0 mm.

3 Experimental Methods

3.1 Imaging Experiments

We acquired high resolution MRI volumes from a 1.5 T Siemens MRI system (Magnetom Symphony, Siemens Medical Systems, Erlangen, Germany). An 8-element phased array body coil was used to ensure coverage of the prostate with a uniform sensitivity. We used a 3D rapid gradient echo sequence (PSIF) designed to acquire the spin-echo component of the steady state response, rather then the free induction decay. The sequence with 9.4/5.0/60 (TR/TE/flip) yielded 160 x 256 x 128 voxels over a 219 x 350 x 192-mm rectangular FOV and 1.4 x 1.4 x 1.5-mm voxels oriented to give the highest resolution for transverse slices. There was over sampling at 31% in the slice direction to reduce aliasing artifacts. The acquisition time was 4.3 min. The sequence gave excellent image contrast for the prostate and its surroundings.

We acquired high resolution MRI volumes from four volunteers S1-S4. For each volunteer, three image volumes were obtained with an imaging session. Each volume was acquired with compatible conditions. Volunteers laid supine with legs flat similar to the position in routine MR scanning. Between volume acquisitions, volunteers got off the MRI table, stretched, and walked around to ensure that they would assume a different position when they laid back on the table. The coil array was centered on the prostate. All images of a volunteer were acquired with the same MRI acquisition parameters. In total, there are 12 pairs of high-resolution MRI volumes for registration.

We used the high-resolution MRI volumes to simulate iMRI images by creating thick slices and adding noise. MR noise is described by the Rician distribution [13], but at reasonably high signal values, the noise is accurately approximated with Gaussian white noise [14]. We added Gaussian noise to the simulated iMRI slice images. Clinically, we typically use an iMRI slice thickness of 4.0 - 6.0 mm. We averaged 3 1.4 mm thick slices to create a 4.2 mm thick slice.

Additionally, we acquired real iMRI images from volunteers S1-S3 using a clinical 0.2 T C-arm open MR scanner (Siemens Open Symphony, Erlangen, Germany). We used a two-dimensional (2D) PSIF sequence with 15.2/7.4/45 (TR/TE/FA) for image slice acquisitions. The iMRI slices were 128x128 with in-plane pixel size of 2.8x2.8 mm and with effective slice thickness of 5 mm.

3.2 Registration Experiments

We used 12 pairs of high-resolution MR volumes to perform registration experiments. For each volume pair, we extracted data from one volume to simulate thick iMRI image slices; and then we registered the simulated image slices to the other volume. We desire an iMRI slice image acquisition method that gives robust, accurate registrations and is relatively insensitive to acquisition parameters. Hence, we performed experiments to determine the dependence on slice orientation (transverse, sagittal and coronal), on slice position relative to the prostate (above, centered, and below) and on image noise from fast imaging techniques.

We also performed SV registration experiments using actual iMRI images. We registered actual iMRI image slices with high-resolution (1.5 T system) MR volumes and

visually evaluated results. For each volunteer S1-S3, there were three high-resolution MR volumes and 30 iMRI image slices giving 90 SV registration experiments, and a total of 270 experiments.

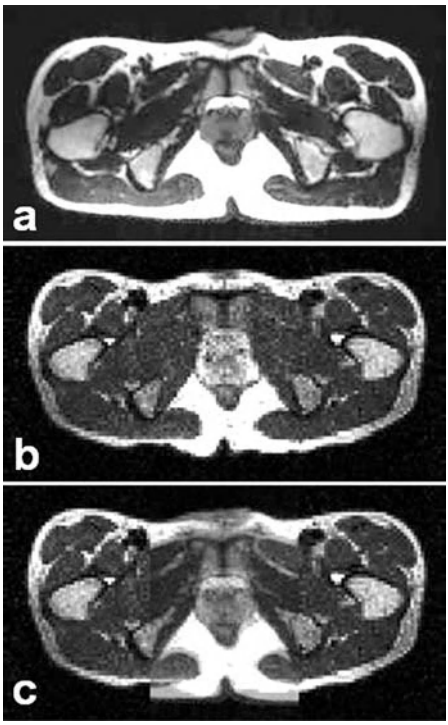


Fig. 3. Prostate images of high resolution MRI (a) and interventional MRI (b). The rectangular region at the center of image (c) is the overlay display of both images. The prostate matches well. Images are from S3.

4 Results

4.1 Simulated Images

Using simulated iMRI images, we determined SV registration results for slices near the prostate in the three standard orthogonal orientations. Comparing to VV, average registration errors were 0.4 mm, 0.5 mm, and 2.6 mm for transverse, coronal and sagittal slices covering the prostate, respectively. Transverse slices worked best because they contained many relatively rigid anatomical structures. Coronal slices worked next best. Sagittal slices gave the largest error because they contained a large portion of the deformable bladder and rectum.

The registration is insensitive to noise. Typical iMRI SNR under clinical conditions is about 25. Even when noise much exceeded this normal situation, registration results were quite good. A 100% success rate was achieved with an acceptance criterion of < 2.0 mm even when SNR was as bad as 10.

4.2 Actual iMRI Images

Registration of actual iMRI image slices with a high-resolution MR volume was successful. The contours overlap and overlay images show that the prostate matches very well. Other visual inspections also demonstrate excellent registration. Note that a single iMRI image was used to produce this registration result.

4.3 Algorithm Implementation

Computation time and registration accuracy are two main factors to consider when choosing interpolation methods. Using tri-linear interpolation, the time for an SV registration was typically about 5 sec on a Pentium IV, 1.8 GHz CPU, with 1Gbyte of memory. When the re-normalized sinc interpolation method was used, the time was $\approx$ 10 min, a duration not acceptable for our application. The algorithm was written in IDL and could probably be made faster in a lower level language such as C. We did not use nearest neighbor because of insufficient accuracy as deduced from its flat peak of the similarity surfaces in Figure 2. A call to the Simplex optimization typically resulted in 50 to 150 similarity evaluations before the tolerance value (0.001) was reached.

5 Discussion and Conclusion

The comparison of similarity surfaces enabled us to design a robust, fast, and accurate registration algorithm for the potential applications of iMRI-guided thermal ablation of the prostate cancer. A single iMRI image slice achieved nearly the same accuracy as obtained from volume-to-volume registration. Since live-time iMRI images are used for guidance and registered images are used for adjunctive information, the registration accuracy is very probably adequate. As compared to a typical SPECT and/or iMRI slice thickness of $\geq$ 3.0 mm, SV registration is quite accurate.

If one were to use functional or high-resolution MR images directly for tumor targeting within the relatively small prostate, the requirements for registration accuracy would be great. However, fused image data will not be used blindly. Rather, these visualizations will be used as a guide. Physicians will always use the live-time iMRI images for needle guidance. With proper visualization tools, physicans should be able to mentally account for any small registration errors. Moreover, the functional images might enable one to find cancer features in the iMRI images.

Finally, we conclude that it is quite feasible to include previously acquired high-resolution MRI and nuclear images into iMRI-guided treatment procedures. We are beginning to explore this application in animal experiments.

Acknowledgements

The algorithm developed in this research was supported by DOD grant DAMD17-02-1-0230 to Baowei Fei, NIH grant R01-CA84433 to David L. Wilson, and ASC grant IRG91-022-06 to Zhenghong Lee. Imaging techniques were developed under the support of NIH grant R33-CA88144 to Jeffrey L. Duerk.

References

1. D.B.Sodee, N.Malguria, P.Faulhaber, M.I.Resnick, J.Albert, and G.Bakale, "Multicenter ProstaScint imaging findings in 2154 patients with prostate cancer," *Urology*, vol. 56, pp. 988-993, 2000.
2. Z.Lee, D.B.Sodee, J.L.Duerk, A.D.Nelson, and M.S.Berridge, "Automatic registration of SPECT-MRI in the pelvis," *Journal of Nuclear Medicine*, vol. 41, pp. 232, 2000.
3. B.W.Fei, J.L.Duerk, and D.L.Wilson, "Automatic 3D Registration for Interventional MRI-Guided Treatment of Prostate Cancer," *Computer Aided Surgery*, vol. 7, pp. 257-267, 2002.
4. B.W.Fei, J.L.Duerk, D.T.Boll, J.S.Lewin, and D.L.Wilson, "Slice to volume registration and its potential application tointerventional MRI guided radiofrequency thermal ablation of prostate cancer," *IEEE Transactions on Medical Imaging*, vol. 22, 2003.
5. N.A.Thacker, A.Jackson, D.Moriarty, and E.Vokurka, "Improved quality of re-sliced MR images using re-normalized sinc interpolation," *Journal of Magnetic Resonance Imaging*, vol. 10, pp. 582-588, 1999.
6. J.V.Hajnal, N.Saeed, E.J.SOAR, A.Oatridge, I.R.Young, and G.Bydder, "A registration and interpolation procedure for subvoxel matching of serially acquired MR images," *Journal of Computer Assisted Tomography*, vol. 19, pp. 289-296, 1995.
7. A.Collignon, F.Maes, D.Delaere, D.Vandermeulen, P.Suetens, and G.Marchal, "Automated multimodality image registration using information theory,"*Information Processing in Medical Imaging: Proc. 14th International Conference (IPMI'95*), Computational Imaging and Vision,pp. 287-298, 1995.
8. F.Maes, A.Collignon, D.Vandermeulen, G.Marchal, and P.Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Transactions on Medical Imaging*, vol. 16, pp. 187-198, 1997.
9. W.H.Press, S.A.Teukolsky, W.T.Vellerling, and B.P.Flannery. *Numerical Recipes in C: The Art of Scientific Computing, Second Edition.* New York: The Press Syndicate of the Cambridge University, 1992.
10. B.W.Fei, A.Wheaton, Z.Lee, J.L.Duerk, and D.L.Wilson, "Automatic MR volume registration and its evaluation for the pelvis and prostate," *Physics in Medicine and Biology*, vol. 47, pp. 823-838, 2002.
11. J.P.W.Pluim, J.B.A.Maintz, and M.A.Viergever, "Image registration by maximization of combined mutual information and gradient information," *IEEE Transactions on Medical Imaging*, vol. 19, pp. 809-814, 2000.
12. J.Nelder and R.A.Mead, "A simplex method for function minimization," *Computer Journal*, vol. 7, pp. 308-313, 1965.
13. A.Macovski, "Noise in MRI," *Magnetic Resonance in Medicine*, vol.36, pp. 494-497, 1996.
14. R.C.Gregg and R.D.Nowak, "Noise removal methods for high resolution MRI," *IEEE Nuclear Science Symposium*, vol. 2, pp. 1117-1121, 1997.

A Protocol for Optimization-Independent Similarity Measure Evaluation

Darko Škerl, Boštjan Likar, Rok Bernard, and Franjo Pernuš

University of Ljubljana, Department of Electrical Engineering
Tržaška 25, 1000 Ljubljana, Slovenia,
{Darko.Skerl,Bostjan.Likar,Rok.Bernard,Franjo.Pernus}
@fe.uni-lj.si

Abstract. Evaluation of a registration method is a complex and application-dependent task. The accuracy and robustness of registration depends on a number of factors, such as image acquisition protocols and parameters, image content, spatial transformation, similarity measure, and optimization. The complex interdependence of these factors makes the assessment of a particular factor on registration difficult even for very specific registration tasks. This paper deals with the evaluation of similarity measures. To reduce the degree of complexity or uncertainty in similarity measure evaluation, we propose an evaluation protocol that enables optimization-independent evaluation. Given the image data and parametric spatial transformation, similarity measure values are sampled equidistantly along random lines in I-dimensional parametric space. The obtained similarity measure profiles are then used to derive statistical estimation of capture range and smoothness of the similarity function and the accuracy, precision, and distinctiveness of its optimum. The proposed protocol is used to evaluate three similarity measures, i.e. mutual information, normalized mutual information, and histogram energy on simulated 2D T1, T2, and PD MR brain volumes. The protocol may be a useful tool, first, for studying the influence of many implementation issues on the similarity function and, second, for selecting the best combination of similarity measure and corresponding optimization method for specific applications.

1 Introduction

Integration of complementary information from images of the same part of the body, of the same or different subjects, acquired by different modalities or by the same acquisition device at different times, is nowadays essential for studying normal and pathological states, diagnosis, and therapy planning, execution, and evaluation [1]. A fundamental task in the integration of image information is registration by which images are brought into the best possible spatial correspondence with respect to each other. The registration methods applied to the field of medical imaging, reviewed in a number of surveys [1-3], may be classified according to the nature of the registration basis and the nature of the transformation. With respect to the first criterion, the methods are classified as point based, segmentation based, or whole image content based methods [3]. In the past a number of robust and accurate registration algorithms have been devised that use the intensities in images alone without any requirement to

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 330–338, 2003.

segment corresponding structures such as points or surfaces. More recently, there has been significant interest in information-theoretic measures, like joint entropy [4,5], mutual information [6,7], and normalized mutual information [7,8]. These measures do not assume any functional relationship between the intensities of the images and are thus applicable to both intermodality and intramodality registrations.

Registrations based on similarity measures simply adjust the parameters of an appropriate spatial transformation model until the similarity measure reaches a local optimum. In an ideal case, the parametrical space, in which the value at each location corresponds to the value of the similarity measure for that transformation estimate would contain a sharp maximum (minimum) with monotonically decreasing (increasing) values with distance away from the optimum position. The limited range of transformations around the optimum for which the value of the measure is a monotonic function of misregistration is called the *capture range*. If the capture range is small, an optimization algorithm will most probably converge to a wrong optimum. Some of these optima may be very small, caused by either interpolation artifacts and/or by a local good match between pixel (voxel) intensities. It is desirable to have information on the similarity function prior to registration, but even in the simple case of rigid registration of 2D images, which requires optimization of three parameters, the parametrical space is too large to analyze the similarity function at every point of the space. Limited information may, however, be obtained if images for which correct registration is known are registered from a large number of starting estimates [9,10] or if one image is systematically translated, usually only in the X and Y directions, and/or rotated from the registered position and the value of the similarity measure is defined for each transformation [7, 11,12]. In this paper we present a more thorough evaluation, excluding optimization, of the similarity function, by which its capture range and smoothness, and the accuracy, precision, and distinctiveness of its optimum are obtained for a certain set of test images. To the best of our knowledge, no protocols exist to evaluate a similarity measure in this way. With this protocol we have evaluated the mutual information, normalized mutual information, and the histogram energy similarity measures. Simulated 2D MR brain data are used for evaluation [13].

2 Methods and Materials

2.1 Similarity Profiles in Normalized Parametrical Space

Parametrical space, characterized by the parameters of the spatial transformation that is supposed to bring two images into correspondence, is first normalized so that equal changes of parameters will have approximately equal impact on the transformation magnitude. This can be achieved either experimentally by analyzing the effects of individual parameters on transformation, or analytically by defining a certain metric on transformation magnitude and normalizing the parameters accordingly. If a parametrical space is normalized, Euclidean metrics may be used to determine distances from the position at which the images are best registered ("gold standard") and where the similarity function has its optimum. Let the position where the images are registered be at the origin of the I-dimensional parametrical space and let SM(X) be the value of a similarity measure at point X; $X=[x_1,\ldots,x_I]$ in this space. Similarity measure values $SM(X_{n,m})$ are defined along N lines and M points evenly spaced on each line.

Each of the N lines is defined by a randomly selected starting position $X_{n,0}$ at a distance R from the origin and its mirror point $X_{n,M}$; $X_{n,M}=-X_{n,0}$. All starting points are thus uniformly distributed on the hyper-sphere with radius R:

$$R = \left\| X_{n,0} \right\|, \tag{1}$$

All N similarity profiles are normalized to the interval [0,1] according to the minimal and maximal values of N profiles. Figure 1 shows a similarity measure profile $SM(X_{n,m})$; m=1,…,M, along the n^{th} line.

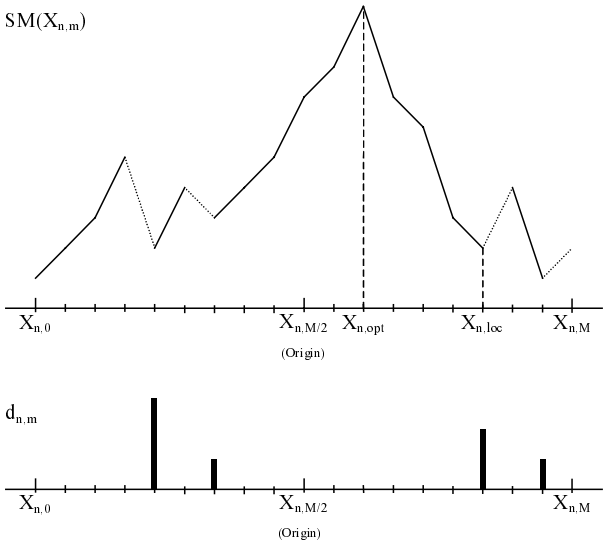


Fig. 1. A typical similarity profile $SM(X_{n,m})$ along n^{th} line with position of global optimum $X_{n,opt}$ and its closest local minimum $X_{n,loc}$ and corresponding positive gradients $d_{n,m}$ (looking from $X_{n,opt}$ outwards).

2.2 Evaluation Criteria

Let $X_{n,opt}$ be the position and $SM(X_{n,opt})$ the value of the global optimum of the similarity measure along line n:

$$\text{opt} = \arg\max_{m}\left(SM(X_{n,m})\right), \tag{2}$$

and let $X_{n,loc}$ be the position of the local minimum closest to $X_{n,opt}$ (Fig. 1).

The *capture range* (CR) of a similarity measure is defined as the minimal of N capturing ranges, estimated as:

$$CR = \min_{n}\left(CR_n\right); \quad CR_n = \left\| X_{n,opt} - X_{n,loc} \right\|, \tag{3}$$

Besides the extent of the capture range it is important to know how the similarity function behaves outside the capture range. This information has been assessed by the

smoothness criterion S, which is the average positive (looking from $X_{n,opt}$ outwards) gradient of all similarity profiles:

$$S = \frac{1}{M}\sum_{m=1}^{M} D_m; \quad D_m = \frac{1}{N}\sum_{n=1}^{N} d_{n,m,poz}, \tag{4}$$

where D_m is the average of N gradients at position m. A gradient $d_{n,m}$ is defined as:

$$d_{n,m} = \begin{cases} pos(SM(X_{n,m}) - SM(X_{n,m+1})) & \text{if } m < opt \\ pos(SM(X_{n,m+1}) - SM(X_{n,m})) & \text{if } m > opt \end{cases}, \tag{5}$$

Only positive gradients (pos) of the similarity function are taken into account as they lead to local optima, away from the global optimum.

Accuracy is defined as the standard deviation σ_a of distances $\| X_{n,opt} \|$:

$$\sigma_a = \sqrt{\frac{1}{N}\sum_{n=1}^{N} \left\| X_{n,opt} \right\|^2}. \tag{6}$$

Precision is defined as the standard deviation σ_p:

$$\sigma_p = \sqrt{\frac{1}{N}\sum_{n=1}^{N} \left\| X_{n,opt} - \overline{X}_{n,opt} \right\|^2}; \quad \overline{X}_{n,opt} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} \left\| X_{n,opt} \right\|}. \tag{7}$$

The difference between accuracy and precision is that accuracy is defined as the standard deviation of N distances between locations of global optima and "gold standard", while precision is defined as the standard deviation of distances between locations of global optima and the average global optimum. The value of precision is always smaller or equal to the value of accuracy.

It is also important to know how the similarity function behaves close to the optimum. When approaching the optimum it may rise steeply and then fall very quickly as well. Or, it can rise slowly and fall slowly after the optimum. We assess this information by the measure of *distinctiveness* (DO), which is the average change of similarity measure value near global optimum. Distinctiveness DO is defined as a function of distance (step *s*=1,2,3,…) from the optimum:

$$DO(s) = \frac{1}{2sN}\sum_{n=1}^{N} 2SM(X_{n,opt}) - SM(X_{n,opt-s}) - SM(X_{n,opt+s}), \tag{8}$$

The function of distinctiveness DO(*s*) provides an estimation of the uncertainty of defining the exact location of the optimum position. Besides, the function DO(*s*) has another practical value in selecting the stopping criteria of an optimization method. These are commonly specified as the minimal change of similarity measure, which is hard to select as it is not related to the physical space, i.e. to the corresponding minimal change of transformation that is related to the expected registration precision. Using DO(*s*) one may easily specify the expected registration precision via *s* and determine corresponding average change of similarity measure that in turn defines corresponding stopping criterion.

2.3 Similarity Measures and Test Images

We have implemented and evaluated three similarity measures: 1) the sum of squared probabilities of the joint histogram (E), mutual information (MI), and normalized mutual information (NMI).

Simulated data were derived from the McGill full anatomical MR brain model images, 181x217x181, 1x1x1-mm voxels [13]. We have evaluated the similarity measures on 4 sets of corresponding slices (74, 80, 90 and 99) of T1, T2 and PD MRI volumes without added noise and intensity inhomogeneities. Figure 1 shows one set of corresponding T1, T2, and PD slices of the brain.

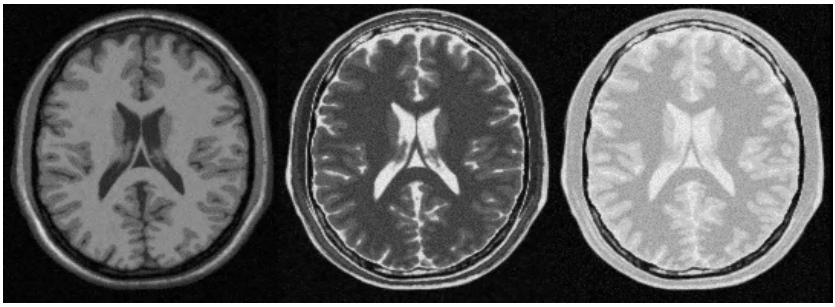


Fig. 2. Corresponding simulated T1, T2, and PD MR brain images.

3 Experiments and Results

For the 2D rigid transformation and for the given test images, which were 181 x 217 pixels large, the following assumption was made to normalize the parametrical space: rotation of the image around its center for 0.1 radians (5.7 degrees) is approximately equivalent to translation for 7 pixels. One unit in the normalized parametrical space thus represents a rotation of 0.1 radians or a translation for 7 mm. In the experiments R was set to 1 when assessing the precision, accuracy and distinctiveness of optimum of a similarity measure and to 15 when the smoothness of a similarity function and its capture range were analyzed. The number of steps M was 60 in all cases and the number of trials N was 150. For each set of T1, T2, and PD 2D images the capture range and smoothness of the similarity function and the accuracy, precision, and distinctiveness of its optimum were assessed for T1-T2, T1-PD, and T2-PD image pairs. The results obtained were very similar for all sets of 2D T1, T2, and PD images tested. In this section we therefore give only the results for the images shown in Fig. 2. Figures 3-5 show curves illustrating the average gradients D_m along N lines (Eq. 4). From these figures the capture range and smoothness of each similarity measure can be deduced. Regardless of the modality, the capturing ranges and smoothness of MI and NMI are almost the same. For all, T1-T2, T1-PD and T2-PD, the capture range for E is very large. Regarding the smoothness of the similarity functions E is the smoothest, followed by NMI and MI.

Tables 1-3 show the capture range, smoothness, accuracy, precision and distinctiveness of global optimum of the three similarity measures tested on T1-T2, T1-PD and T2-PD images, respectively. For all intermodalities tested, the similarity measure E has the largest capture range, MI and NMI are the most precise and accurate in finding the optimum and E is the smoothest and MI has the most distinctive optimum. Evaluations also revealed that mutual information and normalized mutual information behaved very much the same.

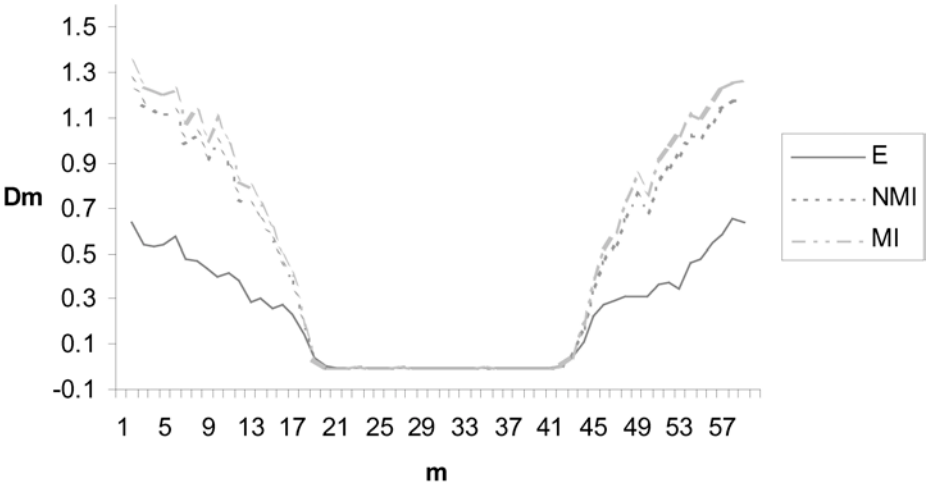


Fig. 3. Average gradients of similarity measures E, MI, and NMI along N lines for T1-T2 images.

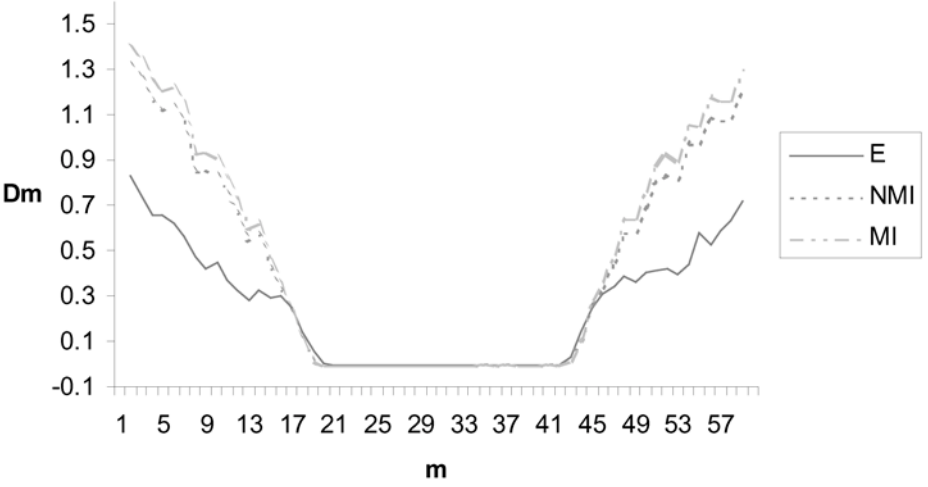


Fig. 4. Average gradients of similarity measures E, MI, and NMI along N lines for T1-PD images.

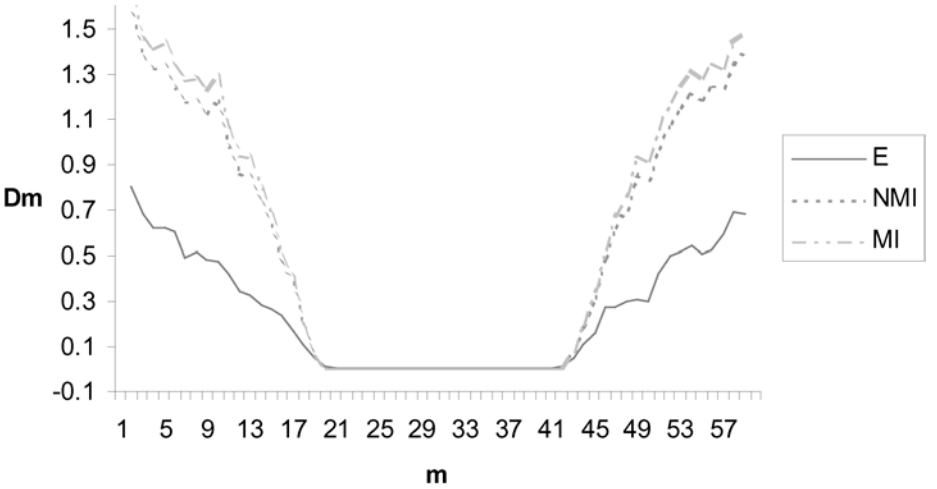


Fig. 5. Average gradients of similarity measures E, MI, and NMI along N lines for T2-PD images.

Table 1. Capture range (CR), smoothness (S), accuracy (σ_a), precision (σ_p) and distinctiveness of global optimum (DO) of the three similarity measures tested on T1-T2 images.

Similarity measure	Criterion				
	σ_a (mm)	σ_p (mm)	CR(mm)	S(1/mm)	DO(1/mm)
E	0.150	0.150	**38.5**	**0.0052**	0.037
NMI	**0.000**	**0.000**	14.0	0.0106	**0.152**
MI	0.019	0.019	14.0	0.0116	0.143

Table 2. Capture range (CR), smoothness (S), accuracy (σ_a), precision (σ_p) and distinctiveness of global optimum (DO) of the three similarity measures tested on T1-PD images.

Similarity measure	Criterion				
	σ_a (mm)	σ_p (mm)	CR(mm)	S(1/mm)	DO(1/mm)
E	0.156	0.155	**38.5**	**0.0057**	0.058
NMI	0.027	0.027	14.0	0.0098	**0.123**
MI	**0.019**	**0.019**	14.0	0.0106	0.116

Table 3. Capture range (CR), smoothness (S), accuracy (σ_a), precision (σ_p) and distinctiveness of global optimum (DO) of the three similarity measures tested on T2-PD images.

Similarity measure	Criterion				
	σ_a (mm)	σ_p (mm)	CR(mm)	S(1/mm)	DO(1/mm)
E	0.149	0.148	35.0	**0.0055**	0.051
NMI	**0.019**	**0.019**	**38.5**	0.0122	**0.115**
MI	**0.019**	**0.019**	**38.5**	0.0132	0.109

Tables 4-6 show the capture range, smoothness, accuracy, precision and distinctiveness of global optimum of the three similarity measures tested on sub-sampled (factor of 2) T1-T2, T1-PD and T2-PD images, respectively. After sub-sampling the rotation of image for 0.1 radians corresponds to 3.5 pixels. The accuracy and precision are smaller then for the original images, which is expected. The capturing ranges are only slightly affected by sub-sampling. All similarity functions are less smooth and the global optimum of each is less distinctive if images are sub-sampled.

Table 4. Capture range (CR), smoothness (S), accuracy (σ_a), precision (σ_p) and distinctiveness of global optimum (DO) of the three similarity measures tested on sub sampled T1-T2 images.

Similarity measure	Criterion				
	σ_a (mm)	σ_p (mm)	CR(mm)	S(1/mm)	DO(1/mm)
E	0.300	0.299	**35.0**	**0.0121**	0.041
NMI	**0.163**	**0.162**	14.0	0.0317	0.038
MI	0.166	0.165	14.0	0.0326	**0.042**

Table 5. Capture range (CR), smoothness (S), accuracy (σ_a), precision (σ_p) and distinctiveness of global optimum (DO) of the three similarity measures tested on sub sampled T1-PD images.

Similarity measure	Criterion				
	σ_a (mm)	σ_p (mm)	CR(mm)	S(1/mm)	DO(1/mm)
E	0.302	0.302	**31.5**	**0.0112**	0.049
NMI	0.174	0.174	14.0	0.0271	**0.067**
MI	**0.172**	**0.172**	14.0	0.0279	0.065

Table 6. Capture range (CR), smoothness (S), accuracy (σ_a), precision (σ_p) and distinctiveness of global optimum (DO) of the three similarity measures tested on sub sampled T2-PD images.

Similarity measure	Criterion				
	σ_a (mm)	σ_p (mm)	CR(mm)	S(1/mm)	DO(1/mm)
E	0.270	0.270	**31.5**	**0.0157**	0.027
NMI	**0.190**	**0.190**	14.0	0.0353	**0.049**
MI	0.192	0.191	14.0	0.0361	0.040

4 Discussion and Conclusions

Similarity measures are usually evaluated according to how well they perform in combination with optimization methods. A similarity measure optimized by different optimizations will most probably give different registration results, therefore we have evaluated similarity measures independently of the optimization process. We proposed a protocol for evaluation, which does not include any optimization. The nature of similarity measure has been assessed by five evaluation criteria, i.e. the capture range and smoothness of the similarity function and the accuracy, precision and distinctiveness of the global optimum. The similarity measures have been tested on original and sub-sampled simulated data derived from the McGill full anatomical MR brain model images. The extension of the protocol to 3D is straightforward but will be

time consuming because more starting points on the sphere would have to be selected and the creation of joint histogram would take more time. Although, the protocol was used on images without added noise and intensity inhomogeneities, it may be used on any monomodal or multimodal images if "gold standard" registration of images is available. Moreover, the protocol can be used to study the influence on a similarity function of many factors, such as sampling, interpolation, histogram binning, and image degradation, such as noise, intensity inhomogeneity, geometrical distortions, partial image overlap, and differences in image content. Results of evaluation may also help to choose the stopping criterion of an optimization process.

References

1. Hill, D.L.G., Batchelor, P.G., Holden, M., Hawkes, D.J.: Medical Image Registration. Physics in Medicine and Biology 46 (2001) R1-R45
2. Hawkes, D.J.: Algorithms for Radiological Image Registration and their Clinical Application. Journal of Anatomy 193 (1998) 347-361
3. Maintz, J.B.A., Viergever, M.A.: A Survey of Medical Image Registration. Medical Image Analysis 2 (1998) 1-36
4. Collignon, A., Maes, F., Delaere, D., Vandermeulen , D., Suetens, P., Marchal, G.: Automated Multi-Modality Image Registration Based on Information Theory. In: Bizais, Y., Barillot, C., Di Paola, R. (eds.): Information Processing in Medical Images. Kluwer Academic, Dordrecht, The Netherlands (1995) 263-274
5. Studholme, C., Hill, D.L.G., Hawkes, D.J.: Multiresolution Voxel Similarity Measures for MR-PET Registration. In: Bizais, Y., Barillot, C., Di Paola, R. (eds.): Information Processing in Medical Images. Kluwer Academic, Dordrecht, The Netherlands (1995) 287-298
6. Wells III, W.M., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R.: Multi- Modal Volume Registration by Maximization of Mutual Information. Medical Image Analysis 1 (1996) 35-51
7. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality Image Registration by Maximization of Mutual Information, IEEE Trans. Medical Imaging 16 (1997) 187-198
8. Studholme, C., Hill, D.L.G., Hawkes, D.J.: An Overlap Invariant Entropy Measure of 3D Medical Image Alignment. Pattern Recognition 32 (1999) 71-86
9. Penney, G.P., Weese, J., Little, J.A. Desmedt, P., Hill, D.L.G., Hawkes, D.J.: A Comparison of Similarity Measures for Use in 2-D-3-D Medical Image Registration. IEEE Trans. Medical Imaging 17 (1998) 586-595
10. Holden, M., Hill, D.L.G., Denton, E.R.E., Jarosz, J.M., Cox, T.C.S., Rohlfing, T., Goodey, J., Hawkes, D.J.: Voxel Similarity Measures for 3-D Serial Brain Image Registration. IEEE Trans. Medical Imaging 19 (2000) 94-102
11. Likar, B., Pernuš, F.: A Hierarchical Approach to Elastic Registration Based on Mutual Information. Image Vision Computing 19 (2001) 33-44
12. Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A.: Mutual Information Matching in Multiresolution Contexts. Image Vision Computing 19 (2001) 45-52
13. Collins, D.L., Zijdenbos, A.P., Kollokian, V., Sled, J.G., Kabani, N.J., Holmes, C.J., Evans, A.C.: Design and Construction of a Realistic Digital Brain Phantom. IEEE Trans. Medical Imaging 17 (1998) 463-468

Medial Node Correspondences towards Automated Registration*

Robert Tamburo[1], Aaron Cois[1], Damion Shelton[2], and George Stetten[1,2]

[1] University of Pittsburgh, Pittsburgh PA 15261, USA
{rjtst2,cacst11}@pitt.edu
[2] Carnegie Mellon University, Pittsburgh, PA 15213, USA
beowulf@cs.cmu.edu,george@stetten.com
http://www.vialab.org

Abstract. Many modern forms of segmentation and registration require manual input making it a tedious and time-consuming process. There have been some successes with automating these methods, but these tend to be unreliable because of inherent variations in anatomical shapes and image quality. It is toward this goal that we have developed an automated method of generating landmarks for registration that will not require supervision or manual initialization. We have chosen medial based image features because they have proven robust against image noise and shape variation, and provide the rotationally invariant properties of dimensionality and scale, which can be used by a unary metric. We introduce a new metric for comparing the geometric relationships between medial features, which overcomes problems introduced by symmetry within a medial feature. With these metrics, we are able to find correspondences between pairs and triplets of features in the two images. We demonstrate these methods on three different datasets. It is envisioned that this system will become the basis for generating medial node models that can be registered between two images.

1 Introduction

For over a century, medical imaging has permitted doctors and other healthcare workers to look beneath the skin's surface to obtain critical diagnostic information. In recent years, computational techniques have been developed to assist clinicians with the analysis of medical images. These techniques include methods for segmentation and registration. Segmentation facilitates the extraction of interesting shapes and the estimation of shape parameters for characterizing disease by classifying regions in the image. Registration is the process of aligning two or more images, valuable for superimposing structures from different imaging modalities and conducting comparative studies of anatomy across a population.

Many current forms of registration require manual input, making it a tedious and time-consuming process. There have been some successes with automated

* This research was supported by a contract with the National Library of Medicine as part of the Insight Toolkit (ITK) software consortium. All software was developed and written with ITK.

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 339–348, 2003.

methods, but these tend to be unreliable because of inherent anatomical variation, image noise, and discontinuous object boundaries [1,2,3]. Our approach is to automatically generate landmarks based on medial features, and we are able to find correspondences between these landmarks in two images. It has previously been shown that correspondences between image features can be found based on geometric constraints [4]. Our method is based on geometric constraints using *cliques* of medial features formed from rotationally invariant metrics. In this paper, we will show that similar objects in two images can be identified by sets of medial features and by the geometric relationships within these sets.

2 Background

We begin with a brief description of medialness, core atoms, and our methods to extract medial features is included in this section. For a more detailed description see [5,6]. The *medial axis* on an object, as defined by Blum for binary images, is the locus of centers of circles that are at least bi-tangent to the object boundary and fit entirely within the object [7]. Pizer et al. developed a measure called *medialness* to extend the medial axis to a medial manifold for gray-level images, where a distinct boundary is replaced by a measure called *boundariness* [8]. Many techniques for measuring medialness (including our own) link the aperture of the boundariness measurement to the radius of the medial manifold. A ridge of medialness is called a *core*. Methods using cores have proven robust against image noise and shape variation [5,6,9,10]. There are several ways to locate the core. We use *core atoms* because they are efficient, requiring only a single application to detect an object [5].

2.1 Core Atoms

The first step in the formation of core atoms is the collection of boundary points. Any boundary detection algorithm can be used so long as it delivers gradient magnitude and orientation. We have chosen to use a Differences Of Offset Gaussian (DOOG) gradient detector [11] whose implementation is described in [12]. A core atom is formed from a pair of boundary points that meet the following requirements:

1. The distance between the boundary points is within a specified range. This distance is termed the *scale*, or *diameter*, of the resulting core atom and is defined as

$$\| \mathbf{s}_{1,2} \| = \| \mathbf{b}_2 - \mathbf{b}_1 \|, \tag{1}$$

where $s_{min} \leq \| \mathbf{s}_{1,2} \| \leq s_{max}$ and the expected diameter is between s_{min} and s_{max}. The vector $\mathbf{s}_{1,2}$ indicates the direction from the first boundary point location $\mathbf{b}_1$ to the second boundary point location $\mathbf{b}_2$ and the core atom is said to be "located" at the midpoint between the two boundary points. Figure 1 illustrates a core atom created across an object of intensity I_1 against a background of intensity I_2.

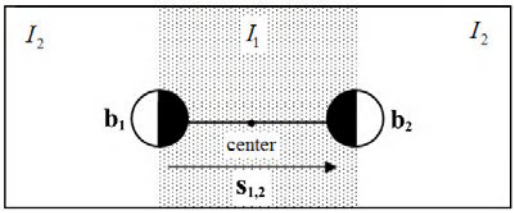


Fig. 1. A core atom consists of two boundary points separated by distance $\| \mathbf{s}_{1,2} \|$ with a center point midway between the boundary locations.

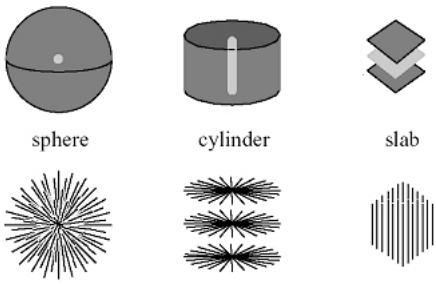


Fig. 2. Top row shows basic shapes in dark gray with corresponding cores in light gray. Below are shown the corresponding basic core atom configurations (from left to right): "koosh-ball," "spokes-of-a-wheel," and "bed-of-nails" (with core atoms shown as line segments).

2. The *face-to-faceness*, as defined in Eq. 2, of the constituent boundaries must be sufficiently close to 1, while still allowing some variation in the relative orientations of those boundaries.

$$F(\mathbf{b}_1, \mathbf{b}_2) = \left| \left(\frac{\mathbf{s}_{1,2}}{\| \mathbf{s}_{1,2} \|} \cdot \hat{\mathbf{n}}_1 \right) \left(\frac{\mathbf{s}_{2,1}}{\| \mathbf{s}_{2,1} \|} \cdot \hat{\mathbf{n}}_2 \right) \right|, \tag{2}$$

where $\hat{\mathbf{n}}_i$, $i = 1, 2$ is the orientation of the i^{th} boundary point (the "$\wedge$" notation denotes a normalized vector).

2.2 Extracting Medial Properties with Core Atoms

Populations of core atoms can be analyzed for a measure of medial *dimensionality*. For 3D data, there are three basic core atom configurations; "koosh-ball," "spokes-of-a-wheel," and "bed-of-nails." The lower half of Fig. 2 illustrates these configurations, where line segments represent core atoms. Above each core atom configuration is the corresponding shape shown in dark gray and the core in light gray.

For a given population of core atoms, the orientation vectors can be statistically processed using eigenanalysis of their covariance matrix, to yield a

measure of dimensionality and an overall estimate of orientation. The eigenvalues $\lambda_1 < \lambda_2 < \lambda_3$ define the dimensionality of the core and the corresponding eigenvectors $\hat{\mathbf{a}}_1$, $\hat{\mathbf{a}}_2$, and $\hat{\mathbf{a}}_3$ define a coordinate system specific to the orientation of the core atom population. The eigenvectors are ordered such that $\hat{\mathbf{a}}_1$ is most orthogonal to the population of core atom vectors and $\hat{\mathbf{a}}_3$ is least orthogonal. The eigenvalues are greater than or equal to 0 and sum to 1. A value of 0 indicates that the corresponding eigenvector is completely orthogonal to every core atom, indicating a direction along the core. This is true for $\hat{\mathbf{a}}_1$ in a cylinder and $\hat{\mathbf{a}}_1$ and $\hat{\mathbf{a}}_2$ in a slab. For the sphere, whose core is a point, no eigenvectors are orthogonal to every core atom. Since the eigenvalues sum to 1, they can be viewed as a system with only 2 independent variables, which define a triangular domain called the *lambda triangle* (shown with results in Fig. 5). The vertices the lambda triangle represent the three basic medial shapes (sphere, cylinder, and slab). All possible sets of eigenvalues are bounded by the lambda triangle.

We sample core atoms first on a regular spatial grid. The core atom centers form a cloud around the true core, because of tolerance in the face-to-faceness. Medial properties measured in samples that are displaced from the true core are prone to distortion. This may cause a misclassification of local dimensionality. Further clustering of core atoms in adjacent samples overcomes this problem, using methods described by Stetten [5]. Each core atom population clustered in this manner is considered a single *medial node*. We define the location of a medial node as the center of mass of the corresponding core atom population, and the scale of the medial node as the average diameter of that population.

3 Methods

Finding correspondences between medial nodes in two images begins with the *unary metric*. Let Ψ_α be the set of all medial nodes in an *N-dimensional* image α. For our applications thus far $N = 3$, but the approach applies to any $N > 1$. The images are designated A and B, namely $\alpha = A$ or B.

3.1 Medial Node Unary Metric

Given that core atoms have been constructed and clustered into nodes, and the dimensionality, scale, and orientation of each node has been determined, the *unary metric* can be calculated between each node in Ψ_A and each node in Ψ_B. The unary metric quantifies the difference between individual medial nodes, independent of orientation, by their medial dimensionality and scale. Recall that each medial node has eigenvalues $\lambda_i, i = 1, 2, \ldots N$ and a scale σ. From these quantities, the unary metric $\mu_{j,k}$ of node j in image A and node k in image B is calculated by

$$\mu_{j,k} = \left(\Sigma_{i=1}^{N-1} \left(\lambda_{j,i} - \lambda_{k,i}\right)^2 \right) \left(\left| \frac{\sigma_j - \sigma_k}{\sigma_j + \sigma_k} \right| \right), \tag{3}$$

where $\lambda_{j,i}$ and $\lambda_{k,i}$ are the i^{th} eigenvalues for node j and k respectively, and σ_j and σ_k are the corresponding scales. Only the first $N-1$ eigenvalues are needed because only that many are independent.

The unary metric has a value between 0 and 1, where 0 indicates identical nodes. All of the metrics found between Ψ_A and Ψ_B are stored in a *unary correspondence matrix* This matrix is $M_A \times M_B$ in size, where M_A and M_B are the numbers of medial nodes in image A and B respectively. The unary correspondence matrix thus stores the similarity between every node in image A and every node in image B.

3.2 Medial Node Signature Metric

In this section, we will define the *signature metric*, to be used in conjunction with the unary metric. We will use the signature metric to find correspondences between similar pairs of medial nodes, also called *2-cliques*, in two images. This metric produces two unique *signatures* for each pair of medial nodes in an image, one signature from the perspective of each node in the pair. The signature metric is used to find corresponding pairs in another image.

Consider a pair of medial nodes in an image. Given the eigenvectors $\hat{\mathbf{a}}_i$, $i = 1, 2, \ldots N$ of node 1, the relative location of node 2 can be expressed in the coordinate system of node 1 as

$$d_j = \sum_{i=1}^{N} v_i \left(\hat{\mathbf{a}}_j\right)_i, \tag{4}$$

where $\left(\hat{\mathbf{a}}_j\right)_i$ is the i^{th} component of eigenvector j and $\mathbf{v}$ is the displacement from node 1 to node 2 in the global coordinate system. We further define $\hat{\mathbf{d}}$ as the normalized vector

$$\hat{\mathbf{d}} = \frac{\mathbf{d}}{\| \mathbf{d} \|}, \tag{5}$$

which expresses the direction to node 2 from node 1 in the coordinate system defined by the eigenvectors of node 1. We will refer to the square of the normalized component $\left(\hat{d}_i\right)^2$ as the i^{th} *Normalized Component Squared* (NCS) of $\hat{\mathbf{d}}$.

A metric to express the similarity between node pairs should ignore symmetry in node 1 in terms of the location of node 2 (and vice versa). Such symmetry results only when two of node 1's eigenvalues are approximately equal. The signature for node 1 is defined as the cumulative NCS function, with the NCS for each component of $\hat{\mathbf{d}}$ (along a given eigenvector $\hat{a}_i$) being added at its corresponding eigenvalue,

$$s(\lambda) = \sum_{i=1}^{N} \left(\hat{d}_i\right)^2 u(\lambda_i), \tag{6}$$

where $u(\cdot)$ is the unit step function.

Since the eigenvalues are positive, $s(\lambda) = 0$ for $\lambda < 0$. In effect, $s(\lambda)$ is a series of steps at each of the eigenvalues, the height of each step being equal to the corresponding NCS. If two or more eigenvalues are equal, their steps superimpose into a single NCS in the degenerate subspace defined by the corresponding

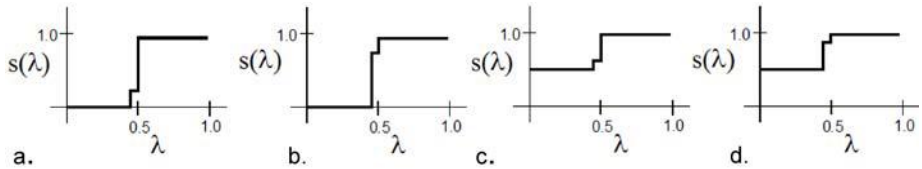


Fig. 3. a) Signature of cylindrical node 1 paired with node 2 at the "equator" of node 1's coordinate system. **b)** Another node pair also with an "equatorial" node 2, but a different orientation of node 1's eigenvectors within the degenerate subspace. **c)** Signature of cylindrical node 1 paired with node 2 at 45° "latitude" of node 1's coordinate system. **d)** Another equivalent node pair, as in Fig 3b.

eigenvectors. By the Pythagorean Theorem, the total step size in such a case is independent of arbitrary rotation within the degenerate subspace. Also by the Pythagorean Theorem, the total of all the steps is $s(1) = 1$.

Typical signatures are shown in Figs. 3a and 3b for a pair of nodes where node 1 is almost perfectly cylindrical (recall that $\lambda_1 = 0, \lambda_2 = 0.5,$ and $\lambda_3 = 0.5$ for a perfectly cylindrical node). In each pair, the second node is at the "equator" of the first node's coordinate system, but there is arbitrary rotation of the eigenvectors within the degenerate subspace around the axis of the cylinder. The two signatures vary along a very short segment of the $\lambda - axis$, and so are very similar. In Figs. 3c and 3d, we show signatures from the same cylindrical node 1, but this time with node 2 located not at the equator, but at 45° "latitude."

Let us look in more detail at the claim that a signature is independent of the global orientation of a node pair. A medial node will produce eigenvectors that depend on orientation with respect to the cardinal coordinate system. However, the eigenvalues are independent of that orientation. When two or more eigenvalues are equal, a degenerate subspace exists that may be expressed by matrix $\mathbf{A}$ whose columns are the corresponding eigenvectors $\mathbf{A} = [\hat{\mathbf{a}}_i\ \hat{\mathbf{a}}_i\ \cdots]$.

Rotating a node pair within the subspace defined by the degenerate eigenvectors of node 1 will redefine the eigenvectors as $\mathbf{E} = [\hat{\mathbf{e}}_i\ \hat{\mathbf{e}}_i\ \cdots] = \mathbf{RA}$, where $\mathbf{R}$ is a pure rotation matrix. The cumulative NCS between node 1 and node 2 will remain unchanged. We demonstrate this for the case of two eigenvalues being equal. Rotation within the degenerate subspace is represented by

$$\begin{aligned} \hat{\mathbf{e}}_i &= \alpha\hat{\mathbf{a}}_i + \beta\hat{\mathbf{e}}_j \\ \hat{\mathbf{e}}_j &= -\beta\hat{\mathbf{a}}_i + \alpha\hat{\mathbf{e}}_j, \end{aligned} \tag{7}$$

where $\alpha^2 + \beta^2 = 1$. The NCS for the two eigenvalues will add to the signature at the same location on the λ axis, namely $\lambda_i = \lambda_j$, because the unit step functions will fuse into a single unit step function, whose height is

$$\left(\hat{d}_i\right)^2 + \left(\hat{d}_j\right)^2. \tag{8}$$

By the Pythagorean Theorem, the quantity $\left(\hat{d}_i\right)^2 + \left(\hat{d}_j\right)^2$ represents the squared length of the hypotenuse, namely, the component of $\hat{\mathbf{v}}$ in the plane

defined by the eigenvectors. This quantity is a constant, independent of which coordinate system was used to calculate it. This can be proven by calculating $\left(\hat{d}_i\right)^2 + \left(\hat{d}_j\right)^2$ in terms of $\mathbf{E}$ and $\mathbf{A}$, and showing them to be equal.

$$\begin{aligned}
\left(\hat{\mathbf{v}} \cdot \hat{\mathbf{e}}_i\right)^2 + \left(\hat{\mathbf{v}} \cdot \hat{\mathbf{e}}_j\right)^2 = \\
\left(\alpha\hat{\mathbf{v}} \cdot \hat{\mathbf{a}}_i + \beta\hat{\mathbf{v}} \cdot \hat{\mathbf{a}}_j\right)^2 + \left(-\beta\hat{\mathbf{v}} \cdot \hat{\mathbf{a}}_i + \alpha\hat{\mathbf{v}} \cdot \hat{\mathbf{a}}_j\right)^2 = \\
\left(\alpha^2 + \beta^2\right)\left[\left(\hat{\mathbf{v}} \cdot \hat{\mathbf{a}}_i\right)^2 + \left(\hat{\mathbf{v}} \cdot \hat{\mathbf{a}}_j\right)^2\right] + \left(2\alpha\beta - 2\alpha\beta\right)\left[\left(\hat{\mathbf{v}} \cdot \hat{\mathbf{a}}_i\right)^2 + \left(\hat{\mathbf{v}} \cdot \hat{\mathbf{a}}_j\right)^2\right] = \\
\left(\hat{\mathbf{v}} \cdot \hat{\mathbf{a}}_i\right)^2 + \left(\hat{\mathbf{v}} \cdot \hat{\mathbf{a}}_j\right)^2. \qquad (9)
\end{aligned}$$

Given two medial node pairs in different images, a *signature metric* of similarity from node 1 to node 2 can now be defined as

$$M_{1,2} = \int_0^1 \left[s_1(\lambda) - s_2(\lambda)\right]^2 d\lambda, \qquad (10)$$

where $0 \leq M_{1,2} \leq 1$. A value of 0 indicates identical signatures. The integral in Equation 10 can be computed efficiently as the area of a series of rectangles by first sorting the union of the two sets of eigenvalues from least to greatest, assigning a new index from $0 \leq h \leq 2N$ to each eigenvalue. The signature metric $M_{1,2}$ is then calculated by going through this set of $2N$ eigenvalues

$$M_{1,2} = \sum_{2N}^{h=1} \kappa\left(\hat{d}_h\right)^2\left(\lambda_h - \lambda_{h-1}\right), \qquad (11)$$

where $\kappa = 1$ for eigenvalues from node 1 and $\kappa = -1$ for eigenvalues from node 2. This yields the same result as the continuous integral in Equation 10.

Cross-image node pairs are said to correspond if (1) the unary metrics for the individual nodes are within tolerance, (2) the distance between nodes for each pair is sufficiently similar, and (3) both signature metrics (from node 1 to node 2 and vice versa) are also within tolerance.

3.3 Medial Node Triplets

Correspondences between cross-image node pairs may not be specific enough. Therefore, we introduce correspondences between medial nodes triplets. Greater specificity of medial node correspondences can be achieved if the geometric configuration between three nodes is considered. The grouping of three medial nodes into a *medial node triplet* or *3-clique* is shown in Fig. 4.

Correspondences between triplets are implied from corresponding node pairs if the difference in angles between them is within a pre-determined threshold. For example, in Fig. 4, suppose node pairs $\left(n_{A,1}, n_{A,2}\right)$ and $\left(n_{A,1}, n_{A,3}\right)$ make up a triplet in image A and corresponding node pairs $\left(n_{B,1}, n_{B,2}\right)$ and $\left(n_{B,1}, n_{B,3}\right)$ make up a triplet in image B. These two triplets correspond if $\mid \theta_A - \theta_B \mid < \tau$, where τ is a pre-determined tolerance. Alternatively, distances L_A and L_B may be used.

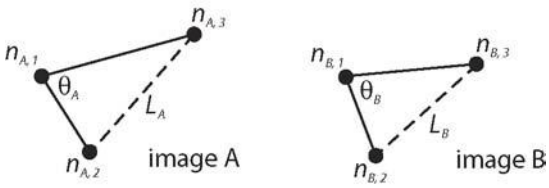


Fig. 4. Triplets from two images. Each vertex is a medial node, where $n_{\alpha,j}$ denotes the j^{th} node of the triplet in image α.

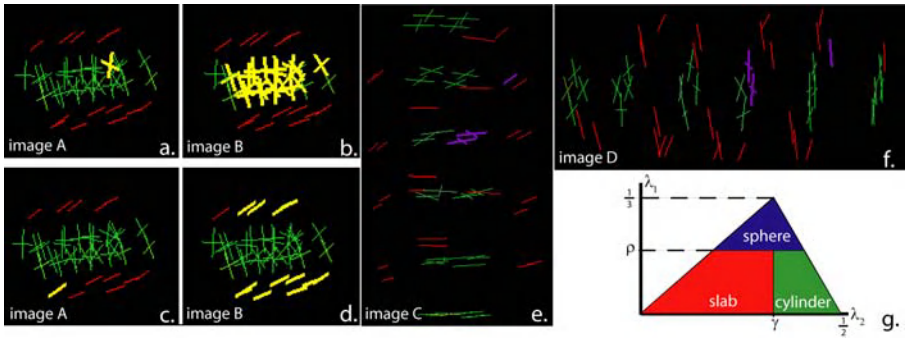


Fig. 5. a) A node within the inner ellipsoid of image A is selected. **b)** The corresponding nodes in image B are highlighted. **c)** A node between the ellipsoids of image A is selected. **d)** The corresponding nodes in image B are highlighted. **e)** A triplet is selected in image C. **f)** The corresponding triplet is highlighted in image D. **g)** The lambda triangle designates dimensionality as a function of eigenvalues λ_1 and λ_2. Arbitrary thresholds placed on λ_1 and λ_2 ($\rho = \frac{1}{5}$ and $\gamma = \frac{1}{3}$) divide the triangle into 3 compartments, which permits dimensionality to be color-coded for visualization purposes.

4 Results

Three sets of data were used to demonstrate the ability of the unary and signature metrics to identify node correspondences. We first used an 8-bit, three-dimensional computer generated dataset consisting of two concentric ellipsoids. Voxels within the inner ellipsoid had an intensity of 255 and voxels between the inner and outer ellipsoids had an intensity of 128.

Correspondences between individual nodes were found between these two images using the unary metric. Figures 5a through 5d show slab-like medial nodes (core atoms that form between the ellipsoids) as red lines and cylindrical medial nodes (core atoms that form across the inner ellipsoid) as green crosshairs, following the color-coding shown in the lambda triangle in Fig. 5g. In Fig. 5a, a node within the inner ellipsoid is selected (yellow crosshair) and the corresponding nodes are shown in Fig. 5b. In Fig. 5c, a node between the ellipsoids is selected (yellow line) and the corresponding nodes are shown in Fig. 5d. As expected only similar nodes are highlighted but there exists more than one cor-

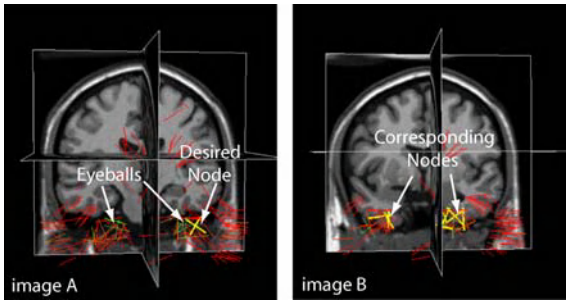


Fig. 6. MRI brain data shows that selecting a node in an eyeball of image A, will highlight corresponding nodes in both eyeballs of image B.

responding node. This demonstrates the utility of the unary metric as well as the need for further specificity using node cliques.

The signature metric was demonstrated on two concentric ellipsoids, the second ellipsoid rotated by 90°. Figure 5e shows a selected node triplet (purple lines) and Fig. 5f shows the only properly corresponding node triplet. Mentally rotating the image superimposes the corresponding nodes.

The unary metric was applied to simulated MRI brain images (See Fig. 6). Image A and image B were randomly generated by the MRI simulator at BrainWeb [13]. Image B was also rigidly deformed by an approximate 15° rotation. The eyeballs are easily identifiable, because their roughly spherical nature is detected as a cylinder. Cylindrical nodes were found in both the eyeballs. Figure 6 shows that a node in one eyeball of image A was selected and corresponding nodes are highlighted in both eyeballs in image B. We are presently applying the signature metric to such image data.

5 Discussion

We have developed a novel method to automatically identify correspondences between medial based image features. The founding metrics were demonstrated to be rotationally invariant and able to locate basic shapes. It is envisioned that this system will become the basis for a means to generate medial node models automatically from a training set, to capture shape variability and locate, as well as measure, objects in a set of images containing similar objects. Useful classification of objects could be achieved based on these correspondences. Potential applications include time series, cross-subject, and multi-modal image analysis.

References

1. C R Maurer, J., Fitzpatrick, J.M.: A review of medical image registration. In Maciunas, R.J., ed.: Interactive Image-Guided Neurosurgery. American Association of Neurological Surgeons, Park Ridge, IL (1993) 17–44

2. Dawant, B.M.: Non-rigid registration of medical images: Purpose and methods, a short survey. In: International Symposium on Biomedical Imaging, Washington D.C., IEEE (2002) 465–468
3. Maintz, J.B.A., Vergever, M.A.: A survey of medial image registration. Medical Image Analysis **2** (1998) 1–37
4. Grimson, W.: Object Recognition by Computer. MIT Press, Cambridge, MA (1990)
5. Stetten, G.: Automated Identification and Measurement of Cardiac Anatomy Via Analysis of Medial Primitives. Doctoral dissertation, University of North Carolina, Chapel Hill (1999)
6. Stetten, G., Pizer, S.M.: Medial-node models to identify and measure objects in real-time 3-d echocardiography. IEEE Transactions on Medical Imaging **18** (1999) 1025–1034
7. Blum, H., Nagel, R.: Shape description using weighted symmetric axis features. Pattern Recognition **10** (1978) 167–180
8. Pizer, S.M., Burbeck, C.A.: Cores as the basis for object vision in medical images. In: Medical Imaging '94: Image Perception, SPIE. Volume 2166. (1994) 191–198
9. Furst, J., Pizer, S.: Marching optimal-parameter ridges: An algorithm to extract shape loci in 3d images. In: Medical Image Computing and Computer-Assisted Intervention. Volume 1496 of Lecture Notes in Computer Science., Cambridge, MA, Springer (1998) 780–787
10. Burbeck, C., Pizer, S.: Object representation by cores: Identifying and representing primitive spatial regions. Vision Research **35** (1995) 1917–1930
11. Young, R.A.: The gaussian derivative theory of spatial vision: Analysis of cortical cell receptive field line-weighting profiles. General Motors Research Publication **GMR-4920** (1985)
12. Tamburo, R.: Gradient-Oriented Boundary Profiles for Shape Analysis Using Medial Features. Master's thesis, University of Pittsburgh (2002)
13. BrainWeb: Simulated brain database (`http://www.bic.mni.mcgill.ca/brainweb/`)

Mouse Brain Spatial Normalization: The Challenge of Sparse Data

Smadar Gefen[1], Oleh Tretiak[2], Louise Bertrand[1], and Jonathan Nissanov[1]

[1] Computer Vision Laboratory for Vertebrate Brain Mapping, Department of Neurobiology and Anatomy, Drexel College of Medicine, Drexel University
2900 Queen Lane Philadelphia, PA 19129-1096
[2] Electrical and Computer Engineering Department, Drexel University
3141 Chestnut St. Philadelphia, PA 19104

Abstract. A three-dimensional surface-based progressive alignment algorithm is proposed to recover nonlinear deformation field. The deformation field is represented with a multi-resolution wavelet expansion and is modeled by the partial differential equations of linear elasticity. We report of its use in spatial normalization of mouse brains reconstructed from sectional material. The wavelet alignment algorithm produced more than threefold improvement in accuracy over an affine (linear) alignment. Its susceptibility to sparse sampling, a problem when the data is derived from tissue sections, was evaluated. Registration accuracy was reduced by only two fold as sampling decreased six-fold.

1 Introduction

Spatial brain normalization is key for intersubject comparison in human neuroimaging studies. Since the imaging modalities typically employed in these studies, namely MRI and PET, yield volumetric data, 3D registration algorithms are suitable – a good thing given that intersubject distortion is in 3D. In contrast, most non-human imaging studies, particularly ones on rodents, rely on ordered sets of sectional material. The advantage of non-invasive 3D imaging with MRI and PET, are greatly offset in these studies by substantially higher-resolution, a vast array of high-specificity contrast agents, and low-cost. The advantage of image fusion [1] and image-based access to neuroinformatics databases [2] that spatial brain nomalization provides is potentially of equal value to its use in human neuroimaging. This is particularly so in the setting of mouse brain studies. The mouse is widely used as a model for neurogenetics in developmental neurobiology and in studies of human disease. The number of animals processed in these is tremendous, many orders of magnitude greater than the number of subjects participating in human neuroimaging studies. The large number of stains and probes used to label the mouse slides produced makes multispectral fusion particularly important.

Spatial normalization in this setting however is problematic. While the interspeciment variability is in 3D, the data collected is an ordered set of 2D images. Registration must either proceed in a 2.5D fashion – correspondence between matching planes established between speciments followed by 2D registration – or 3D reconstruction must precede interspeciement alignment. We have recently reported on performance using the 2.5D approach where plane matching was done manually and 2D registra-

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 349–357, 2003.

tion made use of a novel algorithm [3]. The algorithm estimates the elastic deformation that is typical of biological organs by minimizing the elastic energy as well as by minimizing the sum of squared distances between corresponding surfaces. Wavelets were used to represent the deformation. The wavelet coefficients were estimated separately and progressively. This progressive approach for recovering the deformation signal was shown to be effective for reducing algorithm complexity. The extension of the alignment approach to 3D is described here. An important performance issue in the setting of normalization of images derived from sectional material concerns robustness as a function of sampling. This is of importance due to the fact that consecutive tissue sections are rarely collected. Instead, thin sections, 10-30μm, are typically taken at intervals equal to a few times the section thickness.

Sampling frequency may impact performance in two ways in the 3D approach. Reconstruction accuracy may be compromised and the efficacy of the wavelet-based registration diminished as the frequency is reduced. While reconstruction methods reliant on sequential section-to-section alignment, not surprisingly, fail when intersection interval increases [4], methods that employ external means to guide registration such as fiducials [5], blockface images [6], or surface mapping using structured light [7], would not be affected. With means available to overcome reconstruction error in face of sparse section sampling, the issue remains of how robust interspeciment alignment is with higher intersection intervals. We examine this problem here using the wavelet-base 3D alignment algorithm [8].

2 Deformation Field Representation

The deformation $\mathbf{u}(\mathbf{x})$, that differentiates two corresponding objects, is represented by a separable three-dimensional orthogonal wavelet decomposition. Bringing two objects into alignment requires estimating the wavelet parameter vector, $\mathbf{c}$, that yields: $\mathbf{x}' = \mathbf{x} + \mathbf{u}(\mathbf{x};\mathbf{c})$ where $\mathbf{x}$ is the coordinate vector in the test object space and $\mathbf{x}'$ is the corresponding coordinate vector in the reference object space. The wavelet decomposition of the deformation signal, $\mathbf{u}(\mathbf{x})$, is given by:

$$\mathbf{u}(\mathbf{x})^i = \sum_{\mathbf{k}=0}^{\mathbf{N}_J-1} 2^{-J} c_{J\mathbf{k}}^{i1} \Phi^1(2^{-J}\mathbf{x}-\mathbf{k}) + \sum_{j=R}^{J} 2^{-j} \sum_{s=2}^{8} \sum_{\mathbf{k}=0}^{\mathbf{N}_j-1} c_{j\mathbf{k}}^{is} \Phi^s(2^{-j}\mathbf{x}-\mathbf{k}) \ , \tag{1}$$

where i is an index denoting the three directions, x, y and z, of the vector function $\mathbf{u}(\mathbf{x})$, s is an index denoting the orientation, j is an index denoting the resolution levels, and $\mathbf{k} = (k_x, k_y, k_z)$ is the translational index within the rectangular support: $\mathbf{N}_j \equiv 2^{-j}\mathbf{N} = 2^{-j}(N_x, N_y, N_z)$. Each wavelet coefficient $c_{j\mathbf{k}}^{is}$ is indexed by the direction (i), the orientation (s), the resolution (j), and the spatial location ($\mathbf{k}$) it stands for. The basis functions, $\Phi_{j\mathbf{k}}^s \equiv \Phi^s(2^{-j}\mathbf{x}-\mathbf{k})$, are a tensor product of the one-dimensional *scaling* and *wavelet* functions. The accuracy of this approximation controlled by our choice of the index R that ranges here between 1 and J, where $2^{J+1} = \min(N_x, N_y, N_z)$. Consequently, for example, in the case where

$N_x = N_y = N_z$ the number of parameters, N_c, required to define a deformation in one direction is $N_c \equiv 8 \cdot 2^{3(J-R+1)}$.

In this study we have used semi-orthogonal spline of order three [9]. Semi-orthogonal wavelets are constructed using non-orthogonal scaling functions and non-orthogonal wavelet functions that are orthogonal to each other. Thus, orthogonality is maintained only between spaces that are spanned by either the scaling or wavelet basis functions. Splines are not orthogonal bases; nevertheless, the use of splines is very common due to their symmetry, locality, and good approximation properties [10].

3 Algorithm Description

Generally, solving the problem of nonlinear registration equivalent to solving a nonlinear optimization problem. Estimating the nonlinear deformation that brings together corresponding surfaces requires computing the registration parameters – the wavelet coefficients $c_{j\mathbf{k}}^{is}$ – that minimize a functional $E(\mathbf{c})$. We define a functional: $E(\mathbf{c}) = e(\mathbf{c}) + L(\mathbf{c})$, where $e(\mathbf{c})$ is the sum of squared surface distances and $L(\mathbf{c})$ is the elastic energy. In the sequel the derivations of $e(\mathbf{c})$ and $L(\mathbf{c})$ as a function of the registration parameters and the Marquardt-Levenberg-based optimization progressive procedure are described.

$e(\mathbf{c})$ is the sum of squared surface distances where surface distance is defined as the distance between corresponding points on the test and reference surfaces. Since in practice the corresponding points are unknown, an estimate for this distance is used. Hence, the surface distance, $d_m \equiv \left\| \mathbf{a}(\mathbf{x}_m, \mathbf{c}) - \mathbf{x}'_m \right\|$, is the Euclidean distance between an already mapped point from the test surface, $\mathbf{a}(\mathbf{x}_m \in \mathrm{S}_{\text{test}}, \mathbf{c})$ and the closest point to it on the reference surface, $\mathbf{x}'_m \in S_{ref}$. Similarly, a vector distance is defined as $\mathbf{v}_m \equiv \mathbf{a} - \mathbf{x}'_m$. Therefore, the magnitude being minimized is:

$$e(\mathbf{c}) = \sum_{m=1}^{M} d_m^2 \cong \sum_{m=1}^{M} d(\mathbf{a}(\mathbf{x}_m, \mathbf{c}), \mathbf{x}'_m)^2 \quad . \tag{2}$$

Local dependency of $e(\mathbf{c})$ on the registration parameters, $\mathbf{c}$, is achieved by substituting the Taylor expansion of d_m in the vicinity of $\mathbf{c}_l$ as follows:

$$e(\mathbf{c}) \approx \sum_{m=1}^{M} [d_m(\mathbf{c}_l) + \Delta^T \, grad_c d_m(\mathbf{c}_l)]^2 = e(\mathbf{c}_l) + \Delta^T \mathbf{b} + \Delta^T \mathbf{A} \Delta \quad , \tag{3}$$

where: $\mathbf{b}_p \equiv 2 \sum_{m=1}^{M} d_m(\mathbf{c}_l) \frac{\partial d_m(\mathbf{c}_l)}{\partial c_p}$, $A_{pq} \equiv \sum_{m=1}^{M} \frac{\partial d_m(\mathbf{c}_l)}{\partial c_p} \frac{\partial d_m(\mathbf{c}_l)}{\partial c_q}$, $\Delta \equiv (\mathbf{c} - \mathbf{c}_l)$, and $\frac{\partial d_m(\mathbf{c}_l)}{\partial c_{j\mathbf{k}}^{is}} = \frac{v_m^{\ i}}{d_m} \Phi^s(2^{-j}\mathbf{x}_m - \mathbf{k})$.

The second term, $L(\mathbf{c})$, that is being minimized is the elastic energy:

$$L(\mathbf{c}) = \int [\sum_{p=1}^{3}\sum_{q=1}^{3}\left(\frac{\partial u^p}{\partial x_q}\right)^2 + \left(\frac{\partial u^1}{\partial x_1} + \frac{\partial u^2}{\partial x_2} + \frac{\partial u^3}{\partial x_3}\right)^2] d\mathbf{x} \ . \tag{4}$$

Substituting the signal as represented in (1) results in a linear combination of the integral:

$$h_{pq}^{st}(j,l,\mathbf{k},\mathbf{n}) = \int \frac{\partial \Phi^s(2^{-j}\mathbf{x}-\mathbf{k})}{\partial x_p} \ \frac{\partial \Phi^t(2^{-l}\mathbf{x}-\mathbf{n})}{\partial x_q} d\mathbf{x} \ , \tag{5}$$

where p,q=1,2. In order to simplify the expression for the elastic energy, we assume that the scaling and wavelet functions satisfy a principle we define and refer to here as *quasi orthogonality of order three*. The quasi orthogonality of order three property states that the scaling and wavelet functions are orthogonal to each other and orthogonal to each other's first and second derivatives. The consequence of satisfying this property of triple orthogonality is that $h_{pq}^{st}(j,l,\mathbf{k},\mathbf{n}) = 0$ for terms in the integral containing basis functions from different resolution levels ($j \neq l$) and different orientations ($s \neq t$). In this study, semi-orthogonal wavelet was used in which case the *quasi orthogonality of order three* property approximately satisfied [9].

Let us continue and develop the expression in (4) for one signal component $u^i(j,s) = \sum_{\mathbf{k}=0}^{\mathbf{N}_j-\mathbf{1}} c_{j\mathbf{k}}^{is} \Phi^s(2^{-j}\mathbf{x}-\mathbf{k})$. In this case, (4) is a linear combination of the following terms:

$$\int [\frac{\partial u^i}{\partial x_p}\frac{\partial u^l}{\partial x_q}] d\mathbf{x} = \sum_{\mathbf{n}}\sum_{\mathbf{k}} c_{jk}^{is} c_{jn}^{ls} h_{pq}^{ss}(j,j,\mathbf{k},\mathbf{n}) \equiv \mathbf{c}_j^{is\,T} \mathbf{Q}_{jpq}^s \mathbf{c}_j^{ls} \ . \tag{6}$$

Accordingly, the energy of the signal portion that corresponds to level j and orientation s is:

$$L_j^s(\mathbf{c}_j^s) \equiv \mathbf{c}_j^{s\,T} \mathbf{Q}_j^s \mathbf{c}_j^s \ , \tag{7}$$

$$\text{where } \mathbf{c}_j^s \equiv \left[\mathbf{c}_j^{1s\,T} \quad \mathbf{c}_j^{2s\,T} \quad \mathbf{c}_j^{3s\,T}\right]^T \text{ and}$$

$$\mathbf{Q}_j^s \equiv \begin{bmatrix} 2\mathbf{Q}_{j11}^s + \mathbf{Q}_{j22}^s + \mathbf{Q}_{j33}^s & \mathbf{Q}_{j12}^s & \mathbf{Q}_{j13}^s \\ \mathbf{Q}_{j21}^s & \mathbf{Q}_{j11}^s + 2\mathbf{Q}_{j22}^s + \mathbf{Q}_{j33}^s & \mathbf{Q}_{j23}^s \\ \mathbf{Q}_{j31}^s & \mathbf{Q}_{j32}^s & \mathbf{Q}_{j11}^s + \mathbf{Q}_{j22}^s + 2\mathbf{Q}_{j33}^s \end{bmatrix}$$

Hence the expression for the whole elastic energy is:

$$L(\mathbf{c}) = \mathbf{c}_J^{1\,T} \mathbf{Q}_J^1 \mathbf{c}_J^1 + \sum_{j=R}^{J}\sum_{s=2}^{8} \mathbf{c}_j^{s\,T} \mathbf{Q}_j^s \mathbf{c}_j^s \ . \tag{8}$$

Note though that the matrix $\mathbf{Q}_j^s$ does not depend on the given object's surfaces and deformation. Therefore, it can be computed once offline and then used for the deformation estimation of any input object data. Finally, the elastic energy in (9) is approximated in the vicinity of $\mathbf{c}_l$ using a Taylor series:

$$L(\mathbf{c}) \cong \mathbf{c}_l{}^T \mathbf{Q}\mathbf{c}_l + \Delta^T 2\mathbf{Q}\mathbf{c}_l + \Delta^T \mathbf{Q}\Delta \ . \tag{9}$$

The expressions developed above for the sum of squared surface distances, $e(\mathbf{c})$, and for the elastic energy, $L(\mathbf{c})$, are combined to result in the functional $E(\mathbf{c})$:

$$E(\mathbf{c}) = e(\mathbf{c}) + L(\mathbf{c}) \approx e(\mathbf{c}_l) + \mathbf{c}_l{}^T \mathbf{Q}\mathbf{c}_l + \Delta^T (\mathbf{b} + 2\mathbf{Q}\mathbf{c}_l) + \Delta^T (\mathbf{A} + \mathbf{Q})\Delta \ . \tag{10}$$

Next, our objective is to find Δ, a step in the parameter space, that is in the direction that minimizes $E(\mathbf{c})$. Following the M-L optimization method [11] this can be done by solving:

$$(\mathbf{b} + 2\mathbf{Q}\mathbf{c}_l) + [(\mathbf{A} + \mathbf{Q}) + \alpha\mathbf{I}]\Delta = 0 \ , \tag{11}$$

where $\mathbf{I}$ is the identity matrix and α is a positive scalar. The smaller α is, the closer the M-L is to the Newton optimization method; the larger α is, the closer the M-L is to the gradient (steepest-descent) optimization method.

Each iteration in the M-L optimization algorithm requires computation of the vector $\mathbf{b}$ and the matrix $\mathbf{A}$ and thereby computing the current surface distance d_m. Computing d_m, in turn, requires mapping M points, $\mathbf{x}_m$, along the test surface using the current registration parameters, $\mathbf{c}$. Straightforward computation of the surface distance of a point $\mathbf{x}_m$ is done by searching for the closest point to it among the set of points on the corresponding reference surface. This distance computation, that needs to be done at each iteration and for all the M test points, is computationally expensive especially in the 3D case. Nevertheless, using a distance map, as described in [8], computation time can be reduced significantly.

As mentioned before, our implementation of the optimization algorithm is a progressive one. We take advantage of the multi-resolution representation of the deformation and recover the wavelet parameters that define it in a coarse-to-fine order. We first approximate the deformation with the wavelet decomposition, in (1), including only the lowest resolution level ($j = J$) and the first orientation ($s = 1$). At this stage we find $\hat{\mathbf{c}}_J^1$ that minimizes $E\left(\mathbf{c}_J^1\right)$ following the M-L optimization method. Then we use $\hat{\mathbf{c}}_J^1$ to map the test surface onto a new position that is now closer to the reference surface. The residual deformation that still exists between the test surfaces and the reference surfaces is then recovered using the rest of the wavelet decomposition components that represent the deformation "details" at its different resolutions and different orientations.

4 Experiments

Four mouse brains (genetic strain C57BL/6J) were studied. One brain, referred to here as the *reference* brain, was the atlas brain. The other three brains, referred to here as the *test* brains, were the experimental data. The reference brain was cut into horizontal sections, 17.9μm wide, using a Leica cryopolycut system. All the sections were collected using Tape Windows technique [12]. Sections were dried overnight, stained with cresyl violet (Nissl stain), and then imaged at a resolution of 8 μm/pixel. The test brains were embedded in celloidin [13]. In this case, brains were cut coronally into 30 μm sections, stained with cresyl violet, and then imaged at a resolution of 4.39μm.

In this study we focused on the olfactory bulbs. We looked at twenty structures; the structures' nomenclatures, based on the Franklin and Paxinos atlas [14], are listed in Table 1. The surfaces of these structures were manually delineated right on the images of the sections from each brain. For example, Fig. 1 shows the volume reconstruction of sections from the olfactory bulbs from one of the experimental brains and manually delineated caudal section. Next, the images and the surfaces of each brain were reconstructed into 30 μm resolution isotropic volume as described in [15].

The evaluation of the wavelet-based registration algorithm was done by aligning the test brains onto the reference brain using a subset of the points that lay on test structures' surfaces. We first aligned the brains linearly using surface-based affine method [16] and then recovered the nonlinear deformation everywhere within a support of 256 by 256 by 128 pixels using the spline semi-orthogonal wavelet of order three [9] and including resolution levels $J = 6$ through $R = 4$.

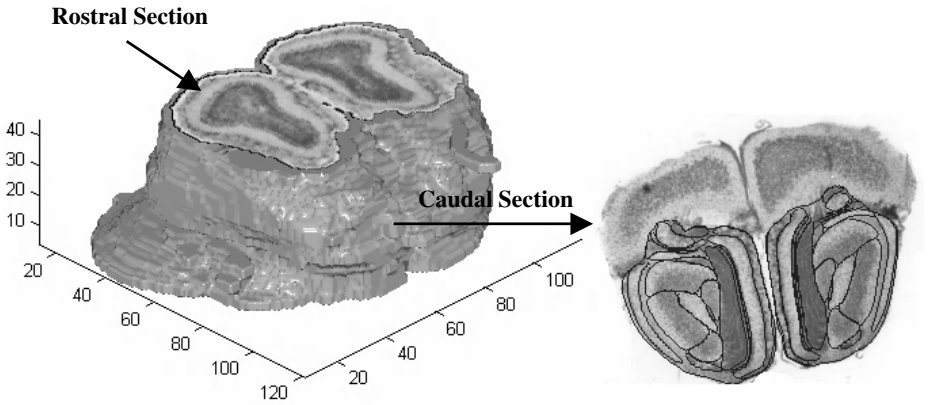


Fig. 1. Volume reconstruction of sections from the olfactory bulbs of experimental data (left) and manually delineated caudal section (right).

The registration errors – average surface distance between corresponding surfaces – were computed across the three test brains. The mean, median and standard deviation of these errors were computed for the data after an affine transform was applied and after the wavelet alignment. Table 1 shows the registration results in the case where points from the test structures' surfaces were sampled at 150 μm interval (5 pixels). It can be seen that on average the wavelet registration method results in accuracy that is more than threefold that of the affine alignment.

In our experiments we have tested the effect of the sampling rate of points from the test structures' surfaces on performance. Being able to register with only limited number of samples from the test dataset is of value since histological images usually involve high volume data. Accordingly, we sub-sampled the surfaces of the test brain's structures at sampling interval ranging between 150μm and 900μm and registered based on these subsets. Fig. 2 shows the registration error as a function of the sampling rate. As expected, increasing the sampling frequency results in a better alignment; when the test points were sampled at 900μm interval (30 pixels) the wavelet alignment average error was 0.45 of the affine average error, while when the test points were sampled at 150μm interval (5 pixels) the wavelet alignment average error was 0.29 of the affine average error. Given that the box that encloses the average test brain is of 3000μm by 2400μm by 1900μm dimension, having twofold improvement in the case where sampling rate was 900μm is an encouraging result.

Table 1. The olfactory bulbs affine and wavelet-based alignment errors using sampling interval of 150 μm for the test dataset.

	Affine Alignment Error			Wavelet-based Alignment Error		
Structures	Mean	Median	STD	Mean	Median	STD
GrO	3.74	2.49	3.34	1.38	0.95	1.33
E/OV	4.35	3.83	2.43	1.56	1.25	1.12
aci	3.96	3.34	3.16	1.38	0.98	1.15
IPl	2.75	2.29	2.11	1.26	0.93	1.01
Mi	3.32	2.04	2.92	1.24	0.94	0.94
EPl	3.45	1.94	3.63	1.10	0.80	1.01
Gl	4.18	2.49	4.40	1.35	0.85	1.39
AOE	7.80	7.68	2.93	1.22	0.87	0.97
vn	8.90	9.00	3.12	3.31	3.28	1.91
EPIA	5.24	5.10	2.89	0.78	0.62	0.52
MiA	6.88	6.72	2.44	1.07	0.92	0.64
AOL	6.56	5.76	4.62	1.27	1.00	0.87
AOM	7.11	6.99	3.52	2.26	1.43	2.12
AOD	7.70	7.57	4.16	1.25	0.95	1.06
dlo	6.40	6.51	2.35	0.91	0.80	0.48
AOV	6.21	5.73	3.69	2.30	1.35	2.33
GrA	4.44	4.16	2.72	1.15	0.88	0.85
VTT	6.72	6.61	3.25	3.23	2.34	2.62
DTT	10.01	9.95	2.72	3.43	3.39	2.10
OB	2.86	1.92	2.71	1.45	1.16	1.08
Average:	**5.63**	**5.11**	**3.16**	**1.65**	**1.28**	**1.27**

Finally, the progressive refinement of the registration accuracy is demonstrated. Improvement in registration accuracy is achieved, clearly, at the price of increasing complexity. Table 2 shows the gain in registration accuracy as lower resolution levels were included in the deformation representation in (1). There is a tradeoff between accuracy and complexity that can be set based on application requirements. The reported algorithm running times are based on algorithm implementation using Matlab environment on 2G Bytes / 1.9GHz Pentium IV PC.

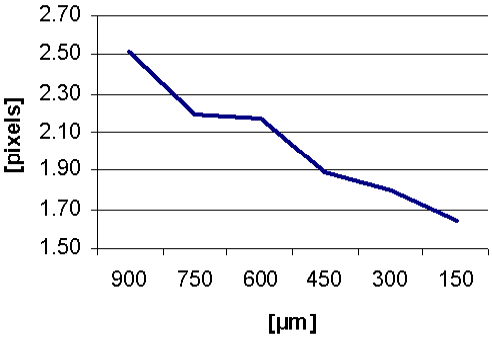


Fig. 2. Registration error as a function of the sampling interval of points from the test structures' surfaces.

Table 2. Registration refinement and resultant increasing algorithm complexity using sampling interval of 300 µm for the test dataset.

Wavelet resolution Levels	Error Mean	Error Median	Error STD	Running Time [min.]	Parameters No.
6	2.46	1.98	1.77	~10	~144
6 , 5	2.11	1.65	1.60	~15	~375
6, 5, 4	1.80	1.41	1.41	~30	~1915

5 Conclusion

Topological matching of histological images is especially challenging because of the intrinsic variability between corresponding regions. Mapping with affine methods, while workable, is not as accurate as can be. This paper presented a mapping method based on wavelet that is on average 3.5 times more accurate. The elastic deformation field was approximated by a 3D separable wavelet decomposition constructed by semi-orthogonal spline of order three. A progressive procedure for the estimation of registration parameters was adopted; the wavelet multi-resolution representation allowed separate handling of the optimization procedure for parameters from different resolution levels.

The robustness and accuracy of the proposed algorithm as a function of sampling rate of surfaces of the test brain's structures was established. It was shown that for a sampling interval that is as low as a third of the test brain average diameter a twofold improvement in alignment can be achieved. The accuracy of the proposed method was measured looking at the reduction in average surface distance between twenty corresponding structures within the olfactory bulbs. It was shown that after affine alignment the wavelet-based method further reduced registration error by a factor that ranges between 2 to 7.

Acknowledgements

This work is supported by the Human Brain Project/Neuroinformatics program funded jointly by the National Institute of Mental Health, the National Institute on Drug Abuse, the National Science Foundation (P20-MH 62009), and NIH training grant NS07440.

References

1. P. Thompson and A. Toga, "Anatomically driven strategies for high-dimensional brain image warpring and pathology detection," in Brain Warping, A. Toga, Ed. New York: Academic Press, 1999, pp. 311-336.
2. T. P. Fox, M. L. Parsons, and J. L. Lancaster, "Beyond the single study: function/location metanalysis in cognitive neuroimaging," Cognitive neuroscience, 1998.
3. S. Gefen, J. O. Tretiak, L. Bertrand, G. D. Rosen, and J. Nissanov, "Surface Alignment of an Elastic Body Using A Multi-resolution Wavelet Representation," revised and resubmitted to IEEE Transaction on Biomedical Engineering, 2003.
4. S. Ardekani, "Inter-Animal Rodent Brain Alignment and Supervised 3D Reconstruction," in School of Biomedical Engineering, Science, and Health Systems. Philadelphia: Drexel University, 2000, pp. 92.
5. A. F. Goldszal, O. J. Tretiak, D. D. Liu, and P. J. Hand, "Multi-modality multi-dimensional analysis of cortical and subcortical plasticity in the rat's vibrissal representation following chronic neonatal deafferentation," , 1994.
6. B. Kim, J. L. Boes, K. A. Frey, and C. R. Meyer, "Mutual Information for Automated Unwarping of Rat Brain Autoradiographs," Neuroimage, vol. 5, pp. 31-40, 1997.
7. J. Nissanov, C. Ozturk, and D. Kozinska, "Brain surface scanning using structured light," presented at SPIE, Philadelphia, PA, 1995.
8. S. Gefen, J. O. Tretiak, and J. Nissanov, "Elastic 3D Alignment of Rat Brain Histological Images," submitted to IEEE Transactions on Medical Imaging, 2003.
9. S. Gefen, "Wavelet-based Nonlinear Multi-dimensional Registration," in Doctoral thesis in Electrical and Computer Engineering. Philadelphia: Drexel University, 2002.
10. G. Strang and T. Nguyen, Wavelets and Filter Banks: Wellesley-Cambridge, 1997.
11. D. W. Marquardt, "An Algorithm for Least-squares Estimation of Nonlinear Parameters," J. Soc. Indust. Appl. Math., vol. 11, pp. 431-441, 1963.
12. J. Nissanov, L. Bertrand, and O. J. Tretiak, "Cryosectioning distortion reduction using tape support," Microsc Res Tech, vol. 53, pp. 239-240, 2001.
13. G. D. Rosen, A. G. Williams, J. A. Capra, M. T. Connolly, B. Cruz, L. Lu, D. C. Airey, K. Kulkarni, and R. W. Williams, "The Mouse Brain Library @ www.mbl.org," presented at Int Mouse Genome Conference, 2000.
14. K. B. J. Franklin and G. Paxinos, The Mouse Brain in Streotaxic Coordinates: Academic Press, 1997.
15. G. Rosen, J. Nissanov, O. Tretiak, K. F. Manly, M. R. Park, and R. W. Williams, "Informatics center for mouse genomics: the dissection of complex traits of the nervous system," Neuroinformatics, Submitted.
16. C. Ozturk, "Alignment Package for Matlab," http://cbis.ece.drexel.edu/ICVC/Align/align.html, 1997.

Towards Physically-Sound Registration Using Object-Specific Properties for Regularization

Yong Zhang, Dmitry B. Goldgof, and Sudeep Sarkar

Department of Computer Science and Engineering
University of South Florida
4202 E. Fowler Avenue
Tampa, FL 33620, USA

Abstract. We study a landmark-based image registration technique that uses elastic model and boundary mapping. Emphasis is given to the reconstruction of heterogeneous material properties and the use of recovered object-specific information to facilitate the registration computation. Tikhonov functional is used to estimate elastic property from landmarks with measurement uncertainty. A preliminary study using a 2D synthetic object indicates that incorporation of actual material properties in the registration method can improve registration accuracy and reduce computational cost.

1 Introduction

Registration is to find a transformation function that maps features in two images to provide complementary information about the object being studied. Registering medical images is particularly challenging because human organs are characterized by nonrigid deformations that can not be easily handled by a global affine model. Various approaches have been proposed to estimate the transformation function that is capable of describing both global and local deformations, among which elastic registration, pioneered by Bajcsy [1], attracts most attentions because of its root in elasticity theory and physical soundness [5,2,3,7]. Elastic registration is often cast as an optimization problem with a generic form:

$$ObjectiveFunction = \phi(\mathbf{u}(\mathbf{x})) = Similarity + Constraints \tag{1}$$

where $\mathbf{u}(\mathbf{x})$ is the displacement vector that specifies the transformation of a spatial point $\mathbf{x}$. Optimal transformation can be found by minimizing the above objective function $\phi(\mathbf{u}(\mathbf{x}))$. Depending upon the context in which registration is performed and the mathematical method being used, the two terms in the above generic formula have been interpreted differently (Table-1).

In landmark-based approaches, it can be written as:

$$\phi(\mathbf{u}) = \sum^{N} |\mathbf{u}(\mathbf{q}) - \mathbf{q}'|^2 + \alpha \|\mathbf{R}\mathbf{u}\|, \tag{2}$$

where $\mathbf{q} = (q_1, ..., q_N)$ and $\mathbf{q}' = (q'_1, ..., q'_N)$ are two sets of corresponding landmarks in two images, $\mathbf{R}$ is a differential operator and α is a weight coefficient.

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 358–366, 2003.

Table 1. Different Views/Interpretations of Generic Registration Form.

View/Framework	Term1	Term2
Image Registration	Similarity	Smoothness
Physical Model	Force	Strain Continuity
Energy Method	External Energy	Internal Energy
Bayesian Method	Likelihood	Prior
Inverse Problem	Residual Norm	Regularization Norm

On the basis of elasticity theory, the first term in equ. (2) can be interpreted as force and the second term accounts for strain energy by having $\mathbf{R} = G\nabla^2 + (\lambda + G)\nabla(\nabla\cdot)$, with G and λ as Lamé constants. To utilize intensity information directly, the force can include other terms such as $\int_V lnP(I_1(\mathbf{u})/I_2)dV$ in the Bayesian framework [5,7], or least squared difference $\int_V |I_1(\mathbf{u}) - I_2|^2 dV$, as being used in [2]. (I_1 and I_2 are intensity function of two images and P is the conditional probability).

If we view image registration as an inverse problem, eq. (2) is a well-posed quadratic functional for a linear system, or a Tikhonov functional in a broader sense, with α as the regularization parameter. If we assume Gaussian noise in landmark correspondence, minimizing $\phi(\mathbf{u})$ is equivalent to estimating $\mathbf{u}$ in Bayesian framework using maximizing *a posterior* method (MAP). For a simple linear system, expectation maximization is equivalent to iterative regularization methods such as Landweber method or iterative Tikhonov method [4].

The choice of α may have a significant impact on the solution $\mathbf{u}$, depending on how severely the original linear system is ill-posed and the magnitude of errors in landmark correspondence. It has been known that global regularization parameter α has a tendency to oversmooth the local discontinuity in the solution space. Terzopoulos [8] suggested the use of piecewise regularization approach to preserve the discontinuity. Other approaches have also been suggested [4]. If we view λ and G as constants throughout the image domain, we can see that λ and G play the role of regularization parameter. If we can obtain spatially variant material properties such as Young's modulus and Poisson's ratio, which are distributed parameters and related to λ and G, then we can solve registration problem by an implicit piecewise regularization scheme.

We propose a registration approach that emphasizes on using heterogeneous material properties to facilitate the registration task. The proposed approach has the following advantages.

1. Registration constraint is posed through actual physical quantities.
2. Using object-specific material properties in elastic model-based registration can improve the registration accuracy.
3. Prior knowledge about material properties of biological tissue can be gathered from standard tensile test.
4. Once reconstructed, object-specific properties can be used to facility registration among other images of the same object. This is especially helpful for multimodality registration.

5. In boundary-based registration scheme, heterogeneous material properties are critical for accurate transformation interpolation.

2 Landmark-Based Registration Scheme

The landmark-based registration approach typically include three steps: (1) global alignment using affine transformation; (2) landmark extraction and correspondence establishment; (3) interpolation of transformation function ($\mathbf{u}(\mathbf{x})$) for all the points in the image, subject to the constraints of landmarks and elastic continuity.

We focus on transformation interpolation by assuming that global affine alignment has been performed and landmark correspondence has been established (with uncertainties). Our registration approach includes four steps: (1) establish landmark correspondence in anchor images; (2) estimate material properties from landmarks in anchor image; (3) establish boundary mapping between regular images; (4) perform registration between regular images through forward modeling synthesis.

We classify images into anchor image and regular image. By "anchor image" we mean that we already have dense landmark correspondence established in the images. The anchor images will only be used for material property reconstruction. Our intention is to perform registration among regular images (or between anchor and regular images), utilizing the material properties recovered from anchor images. We assume that anchor images and regular images are taken for the same object.

The quality of estimated material properties is dependent on the density of landmarks and accuracy of landmark correspondence. Ideally, landmarks should be obtained from modalities that are suited for landmark extraction such as high resolution MRIs.

To register two regular images, we designed a scheme that takes advantage of the predictive ability of forward physical model. We first establish correspondence between landmarks on the boundary of the object and then synthesize the the transformation $\mathbf{u}(\mathbf{x})$ for all points that are inside the object. This scheme will reduce the computation load considerably, since once reconstructed, the material properties can be used to register any images of the same object.

3 Estimating Heterogeneous Properties from Landmarks

Forward Model of Elastic Object. For a linear elastic body, its deformation can be described by the following partial differential equation:

$$(\lambda + G)\nabla(\nabla \cdot \mathbf{u}) + G\nabla^2\mathbf{u} + \mathbf{F} = \rho\frac{\partial^2\mathbf{u}}{\partial t^2} \tag{3}$$

To compute displacement for points on the body, we transform the governing equations into the algebraic equations through finite element discretization. The

resulting transient equation in the matrix form becomes: $\mathbf{M}\ddot{\mathbf{u}} + \mathbf{C}\dot{\mathbf{u}} + \mathbf{K}\mathbf{u} = \mathbf{F}(t)$, where $\mathbf{M}$ is the mass matrix, $\mathbf{C}$ is the damping matrix, $\mathbf{K}$ is the stiffness matrix in which the material properties are embedded, and $\mathbf{F}(t)$ is the force. In static simulation, it becomes $\mathbf{K}\mathbf{u} = \mathbf{F}$. $\mathbf{F}$ can be implemented as Dirichlet condition or Neumann condition.

Tikhonov Regularization. Considering the equation for static case:

$$(\lambda + G)\nabla(\nabla \cdot \mathbf{u}) + G\nabla^2\mathbf{u} + \mathbf{F} = 0, \tag{4}$$

With $\mathbf{E}$ as Young's modulus, we define a differential operator $B(\mathbf{E})$ as:

$$B(\mathbf{E}) = (\lambda + G)\nabla(\nabla \cdot (\cdot)) + G\nabla^2(\cdot), \tag{5}$$

Rearranging the governing equation (4) using $B(\mathbf{E})$, we derive a nonlinear operator equation:

$$B(\mathbf{E})\mathbf{u} + \mathbf{F} = 0, \tag{6}$$

$$A(\mathbf{E}) = -\mathbf{F}B(\mathbf{E})^{-1} = \mathbf{u}. \tag{7}$$

We assume that the nonlinear operator $A : X \to Y$ is continuously Fréchet differentiable in the Hilbert space (X, Y). Given the noisy measurement data, we have an inverse problem of estimating Young's modulus as following:

$$A(\mathbf{E}) = \mathbf{u}. \tag{8}$$

$$\|\mathbf{u}_t - \mathbf{u}^\delta\| \leq \delta, \tag{9}$$

where $\mathbf{u}_t$ is the true displacement, $\mathbf{u}^\delta$ is the corrupted measurement with a noise level of δ.

This inverse problem is very likely ill-posed in Hadamard sense. To overcome the numerical difficulties caused by the ill-posedness, We reformulate the parameter estimation problem as Tikhonov functional in its variational form:

$$T(\mathbf{E}) = \|\mathbf{D}A(\mathbf{E}) - \mathbf{u}^\delta\|^2 + \beta\|\mathbf{W}(\mathbf{E} - \mathbf{E}^*)\|^2, \tag{10}$$

where $\mathbf{D}$ is the projection/interpolation matrix that maps the continuous degrees of freedom to the discrete measurement coordinates, $\mathbf{E}^*$ denotes the prior knowledge about the Young's modulus, β is the regularization parameter and $\mathbf{W}$ is the smoothness matrix (a discretized version of gradient or Laplacian operators).

Minimization. We use the Gauss-Newton method to solve the nonlinear ill-posed inverse problem. In the following derivation, we use $\mathbf{P}$ to denote the generic parameter vector to be estimated, which could be Young's modulus E, Poisson's ratio ν or any other quantities of interest. $\mathbf{d}^\delta$ is the data vector that consists of measured displacement as well as boundary conditions of Neumann type. We also drop the interpolation matrix $\mathbf{D}$, assuming that the output of forward model has been computed on the measurement coordinate.

Given the Tikhonov functional of a nonlinear operator $A(\mathbf{P})$:

$$T(\mathbf{P}) = \|A(\mathbf{P}) - \mathbf{d}^{\delta}\|^2 + \beta\|\mathbf{W}(\mathbf{P} - \mathbf{P}^*)\|^2. \tag{11}$$

We linearize (11) around a local point $\mathbf{P_k}$:

$$L(\mathbf{P}) = \|A(\mathbf{P_k}) + J(\mathbf{P_k})(\mathbf{P} - \mathbf{P_k}) - \mathbf{d}^{\delta}\|^2 + \beta\|\mathbf{W}(\mathbf{P} - \mathbf{P}^*)\|^2, \tag{12}$$

where $J(\mathbf{P})$ is the Fréchet derivative and $L()$ denotes linearized functional. We then minimize $L(\mathbf{P})$ through the vanishing gradient condition:

$$g(\mathbf{P}) = \nabla L(\mathbf{P}) = 0, \tag{13}$$

$$J^T(\mathbf{P_k})(A(\mathbf{P_k}) + J(\mathbf{P_k})(\mathbf{P} - \mathbf{P_k}) - \mathbf{d}^{\delta}) + \beta\mathbf{W^T W}(\mathbf{P} - \mathbf{P}^*) = 0. \tag{14}$$

Rearranging (14), we get iterative fixed point formula of estimating $\mathbf{P}$:

$$\mathbf{P_{k+1}} = \mathbf{P_k} + \triangle\mathbf{P}, \tag{15}$$

$$\triangle\mathbf{P} = [J^T(\mathbf{P_k})J(\mathbf{P_k}) + \beta\mathbf{W^T W}]^{-1}[J^T(\mathbf{P_k})(\mathbf{d}^{\delta} - A(\mathbf{P_k})) - \beta\mathbf{W^T W}(\mathbf{P_k} - \mathbf{P}^*)] \tag{16}$$

Eq. (16) is the normal equation of following linear system:

$$\begin{bmatrix} J(\mathbf{P_k}) \\ \sqrt{\beta}\mathbf{W} \end{bmatrix} \triangle\mathbf{P} = \begin{bmatrix} \mathbf{d}^{\delta} - A(\mathbf{P_k}) \\ -\sqrt{\beta}\mathbf{W}(\mathbf{P_k} - \mathbf{P}^*) \end{bmatrix}. \tag{17}$$

So, we can compute the parameter increment $\triangle\mathbf{P}$ by solving a linear system (17) using the Conjugate Gradient method.

4 Registration through Forward Synthesis

The drawback of registration methods that rely on large number of landmarks is that the landmarks may not be available everywhere inside the images. Davatzikos *et al* [3] proposed a boundary-based approach where the full-image registration is inferred from boundary mapping. This is essentially a Dirichlet boundary value problem, i.e. a forward modeling. The basic idea is that, given appropriate material properties, we can synthesize the deformation insides the object from boundary observations. In the following discussion, we focus on the synthesis (interpolation) process by assuming that boundary mapping has been done. Using finite element method, we implement the synthesis scheme with two steps: (1) computing transformation $\mathbf{u}(\mathbf{x})$ for all nodal points that are inside the object by driving the forward finite element model with Dirichlet boundary conditions; (2) interpolate $\mathbf{u}(\mathbf{x})$ for points insides the element from nodal solutions.

Using a static model, we express the first step of forward synthesis in two equations:

$$\mathbf{Ku} = \mathbf{d} \tag{18}$$

$$\mathbf{y'_n} = \mathbf{y_n} + \mathbf{u}(\mathbf{y_n}) \tag{19}$$

where $\mathbf{K}$ is stiffness matrix that is composed of recovered heterogeneous material properties. $\mathbf{d}$ is the boundary displacement vector determined from boundary mapping, and $\mathbf{u}$ is the displacement for all the finite element nodes. $\mathbf{y'_n}$ and $\mathbf{y_n}$ are locations of the finite element nodes before and after forward modeling, $\mathbf{u}(\mathbf{y_n})$ is computed nodal displacement (transformation).

In second step, we interpolate $\mathbf{u}(\mathbf{x})$ for points inside the element from nodal solution:

$$\mathbf{u_e} = \mathbf{Su} \tag{20}$$

$$\mathbf{y'_e} = \mathbf{y_e} + \mathbf{u_e} \tag{21}$$

where $\mathbf{u_e}$ represents the displacement inside the elements, $\mathbf{u}$ is the nodal displacement obtained from the previous step, and $\mathbf{S}$ is the shape function (basis function) used to interpolate $\mathbf{u_e}$ from $\mathbf{u}$. $\mathbf{y'_e}$ and $\mathbf{y_e}$ are locations of points inside an element before and after deformation.

Although quantitative evaluation of the impact of property heterogeneity on the quality of model-based elastic registration has not be found, the importance of using object-specific material properties in elastic registration is widely recognized and study using variable properties based on anatomical structure has been reported [6]. Our experiment discussed in the next section demonstrates that using object-specific heterogeneous property is critical to the success of boundary-based registration.

5 Experiments

Images. The object used in the experiment is an elastic bandage with known Young's modulus (8 kPa). A small piece of less elastic tape (37 kPa) was attached to the bandage to mimic property abnormality (heterogeneity). A 6x6 rectangle grid was printed on the bandage to help establishing landmark correspondence. Three intensity images were acquired using Minolta range camera while the bandage was stretched laterally to generate nonrigid deformation (Figure 1). The first two frames were used as anchor images to estimate the Young's modulus, and the first and the third frames were used to demonstrate the proposed registration scheme.

Estimated Material Properties. We used the intersections of the grid as landmarks in two anchor images ((Figure 1 (d),(e)). For illustration purpose, we build the finite element model by designing a mesh that is the same as the grid printed on the bandage, using landmarks as the nodes. This simple 2D model contains 49 nodes and 36 rectangle elements. The reconstructed Young's modulus was shown in Figure 1 (f). The abnormality was successfully identified with an average Young's modulus of 29 kPa for area that is covered by the tape, and 12 kPa for the bandage.

Registration. As the first step, we established correspondence for the boundary nodes between frame-1 and frame-3 (Figure 2 (a),(b)). The displacements

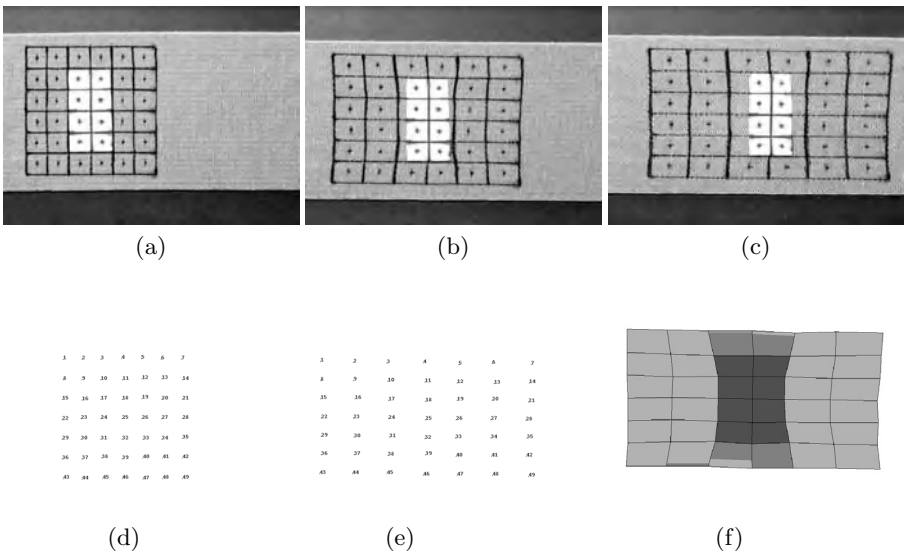


Fig. 1. Reconstruction of material properties using landmarks in anchor images. (a) Anchor image 1. (b) Anchor image 2. (c) Regular image to be registered with anchor image 1. (d) Landmarks in anchor image 1. (e) Landmarks in anchor image 2. (f) Reconstructed Young's modulus, dark color (blue) indicates material abnormality (high Young's modulus value).

of the boundary nodes were then used to specify the Dirichlet condition of the finite element model. In the interpolation step, using this boundary data and recovered material properties, we computed the displacements for all the nodes inside the model, and then synthesized the displacement for points in the middle of rectangle element. To study the influence of material heterogeneity on registration accuracy, we performed two experiments, one with a uniform Young's modulus of 8 Kpa throughout the model, and another one with recovered heterogeneous distribution. The synthesized position for all the middle points of each element in frame-3 were shown as squares in Figure 2 (c),(d).

It is obvious that a simple assumption of uniform material properties causes large registration errors, particularly in the area adjacent to property discontinuity. If we compute the error as ratio between the magnitude of mismatch and true displacement of the middle points, the average errors are about 29% around the discontinuity. In real applications, the variation of property values among different tissues could be much larger than the values used in this experiment (bones, ligament, fat and fluid etc.), a registration approach without considering property heterogeneity could cause significant errors.

6 Conclusions

We presented a registration approach that uses the object-specific material properties to constrain the computation of transformation function. We utilized the

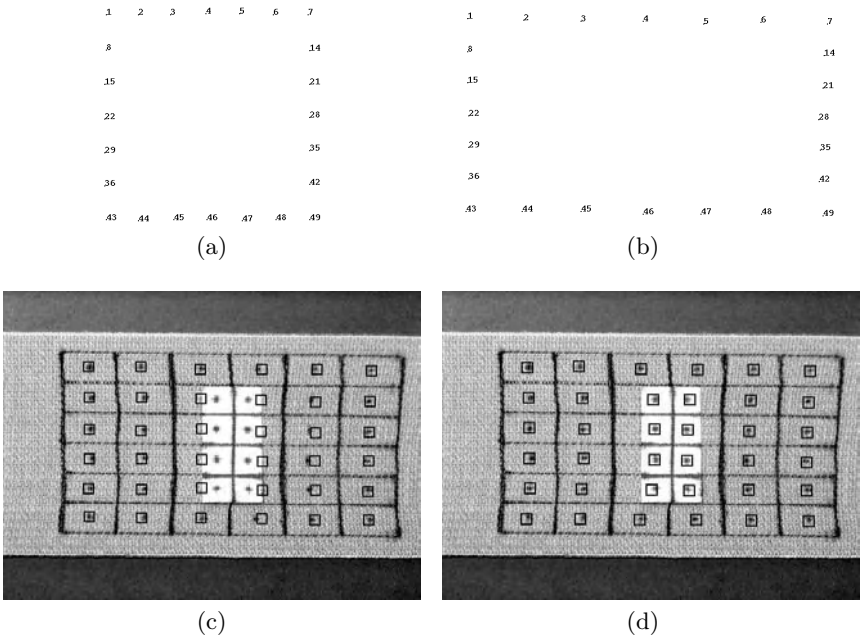


Fig. 2. Boundary-based registration using recovered heterogeneous material properties. (a) Boundary landmarks in anchor image 1. (b) Boundary landmarks in regular image (frame 3). (c) Registration results using assumed uniform Young's modulus. (d) Registration results using recovered heterogeneous Young's modulus. The black squares indicate the synthesized positions for the middle point of each element. Note the mismatch between synthesized positions and real positions in (c).

Tikhonov regularization technique to estimate the heterogeneous material properties from two anchor images that contain landmark correspondence information. We then used the recovered properties to facilitate the boundary-based registration. The advantage of the proposed approach is its physical soundness, because regularization constraint is posed through actual physical properties of the object being registered. More importantly, the recovered material properties can be used to register any images of the same object.

References

1. Bajcsy, R., and Kovacic, S.: Multiresolution elastic matching. Computer Vision, Graphics and Image Processing **46** (1989) 1–21
2. Christensen, G.E., Joshi, S.C., Miller, M.I.: Volumetric transformation of brain anatomy. IEEE Transactions on Medical Imaging **16** (1997) 864–877
3. Davatzikos, C., Prince, J.L., Bryan, R.N.: Image registration based on boundary mapping. IEEE Transactions on Medical Imaging **15** (1996) 112–115
4. Engl, H.W., Hanke, M., Neubauer, A.: Regularization of Inverse Problems. Kluwer Academic Publishers (1996)

5. Gee, J., Barillot, C., Briquer, L.L., Haynor, D., Bajcsy, R.: Matching structural images of the human brain using statistical and geometrical image features. SPIE Visualization in Biomedical Computing **2359** (1994) 191-204
6. Hagemann, A., Rohr, K., Stiehl, S.H., Spetzger, U., Gilsbach, J.M.: Biomechanical modeling of the human head for physically based, nonrigid image registration. IEEE Transactions on Medical Imaging **18** (1999) 875–884
7. Miller, M.I., Christensen, G.E., Amit, Y., Grenander, U.: Mathematical textbook of deformable neuroanatomies. Proceedings of the National Academy of Sciences **90**(24) (1993) 11944–11948.
8. Terzopoulos, D.: Regularization of visual problems involving discontinuities. IEEE Transactions on Pattern Analysis and Machine Intelligence **8** (1986) 413–424

Characterizing Shape Differences between Phantom Image Populations via Multivariate Statistical Analysis of Inverse Consistent Transformations

Dinesh Kumar[1], Xiujuan Geng[1], Gary E. Christensen[1], and Michael W. Vannier[2]

[1] Department of Electrical and Computer Engineering
The University of Iowa, Iowa City, IA, 52242
{dinesh-kumar,xiujuan-geng,gary-christensen}@uiowa.edu
[2] Department of Radiology
The University of Iowa, Iowa City, IA, 52242
michael-vannier@uiowa.edu

Abstract. This paper presents an improved method for detecting statistically significant shape differences between image populations based on the multivariate Hotelling's T^2 test applied directly to image transformations. Performance of the method was evaluated using two phantom populations each consisting of 30 2D images with known average shape and known shape variability. Inverse-consistent linear-elastic image registration (ICLEIR) was used to construct a deformable template average image for both populations. The average image for the "normal" phantom population was used as the reference coordinate system to estimate ICLEIR correspondence transformations from the reference image to each population image. Following the work of Thirion et al., a multivariate two population Hotelling's T^2 test was performed on the displacement fields of these transformations at each voxel location in the reference coordinate system. We show that adding a conditioning constant ϵ to the singular values of the sample covariance matrices used in the Hotelling's T^2 test reduces the false-positive rate. Furthermore, it is shown that adjusting the value of ϵ focuses the statistical test response to the region of known shape differences present in the phantom image populations. Although limited, the phantom results presented in this paper provide baseline information for interpreting future results generated from real 3D medical images.

1 Introduction

Statistical methods are needed to detect and characterize subtle shape differences between anatomical populations in the presence of normal anatomical shape variability. Applications include detecting differences between normal and abnormal populations to improve diagnosis, treatment, and post treatment evaluation and detecting differences between normal populations due to gender, handedness, and age. In general, the location and magnitude of shape differences between real anatomical populations cannot be predicted in advance. Therefore it is important to use a validated method to have confidence in the produced results. Unfortunately, validation of methods used to detect and characterize population shape differences remains a difficult task since there is no "gold" standard to compare with. In this paper, we present an improved method for detecting statistically significant shape differences between image populations and validate

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 367–376, 2003.

the method using 2D phantom images with known shape and shape variability. These phantom results will be used to interpret future results produced from real 3D medical images.

The multivariate two population Hotelling's T^2 test is one of the most important statistical techniques for detecting significant differences between two populations [1]. The importance of this method lies in the fact that it provides joint multivariate analysis of correlated variables. Recently, the T^2 test has been applied to to describe brain shape variability by Thompson and Toga [2]. In their work, a collection of 18 normal brains globally registered to a reference stereotaxic space were defined as a probabilistic brain atlas. A high-dimensional elastic warping algorithm was used to register a candidate brain image to each brain in this atlas. For each point in the candidate brain, a T^2 distribution was fit to its 18 corresponding point locations in the atlas space. A T^2 test was then performed to determine whether or not the location of the point in the candidate brain was significantly different than its corresponding points in the atlas space.

An alternative approach for studying neuroanatomical shape differences based on a two population t-test is called voxel-based morphometry (VBM) proposed by Ashburner and Friston [3,4,5]. In their approach, gray matter is segmented from brain images and are registered to a reference coordinate system using affine and higher order nonrigid image registration. Statistically significant population shape differences are detected by computing the two population t-test on the intensity of Gaussian filtered gray matter segmentations at each voxel in the coordinate system of the reference data set.

We follow the approach proposed by Thirion et al. [6] in which a two population multivariate T^2 test applied to the displacement fields at each voxel in a reference coordinate system to detect and localize significant shape differences between image populations. The 2D deformation field is computed from a template image to all the images in both populations. Statistical shape differences are detected at the voxel level by computing the two-population Hotelling's T^2 test on each component of the deformation fields at every pixel in the reference coordinate system.

This paper reports on the validation of multivariate shape difference determination in populations using inverse consistent linear elastic image registration (ICLEIR) [7]. In contrast to [6] which presents multivariate T^2 test without showing any multivariate results, this paper presents results comparing two phantom image populations with known shape differences. Furthermore, we show that it is necessary to condition the sample covariance matrices used to compute the T^2 statistic in order to reduce the false positive rate to an acceptable level. Results are presented that show that the multivariate Hotelling's T^2 statistic with sample variance conditioning is able to locate the known shape difference between two phantom image populations. In the future, we plan on using this method to determine the location and magnitude of small asymmetries in brain shapes based on comparison of populations.

The rest of the paper is organized as follows. The next section describes the methods used for ICLEIR registration and generating the population average using the transformations. Next, the Hotelling's multivariate T^2 test as applied to transformations and sample covariance matrix conditioning are described. The results of applying our statistical methods to two populations of 30 phantom images each are described in in the results section. The effects of varying the conditioning constant ϵ are studied as they

relate to focusing the response region to the region with large shape differences. Finally, a summary and conclusions for this study are presented.

2 Methods

This section describes the methods used for generating the average shape of a population by averaging transformations and performing statistical hypothesis testing on the transformations to determine population shape differences.

2.1 Phantom Image Populations

Two populations consisting of 30, 256×256 pixel phantom images were generated for this study (see Fig. 1). The first image population consisted of roughly circular shaped objects centered at (128,128) and the second image population consisted of similar shaped objects with an additional bump at $45\,\mathrm{deg}$. The object outer contour for images in population 1 was defined in polar coordinates as $r = 80 + \rho \sin(20\,\theta + \phi)$ where ρ and ϕ were random variables uniformly distributed in the ranges $(0, 2.4)$ and $(0, 2\pi)$, respectively. Therefore, the mean shape of population 1 is a circle of radius 80 pixels. The second population was created in the same way with an additional elliptical bump with a major and minor axis of 21 and 14 pixels, respectively, inclined at $45\,\mathrm{deg}$. The angle of $45\,\mathrm{deg}$ ensures that the shape differences between the populations would be encoded in both the x- and y-displacement fields. The average center location of the bump was on a $45\,\mathrm{deg}$ line at a radial distance of 62 pixels from the center of the circle. The center position of the elliptical bump was modeled as having a random deviation in x and y-directions uniformly distributed in the range of $(-2.4, 2.4)$ pixels. The objects were not corrected for global scale differences in this study since all the objects were centered at (128,128) and had the same expected width and height by construction.

2.2 Inverse-Consistent Linear-Elastic Image Registration

Inverse-consistent linear-elastic image registration (ICLEIR) [7] was used to estimate all correspondence transformations between images in this study. In this approach, the forward transformation $h_{i,j}$ from template image T_i to target image T_j and the reverse transformation $h_{j,i}$ from T_j to T_i are estimated jointly while minimizing the inverse consistency error. A brief summary of this registration method is presented.

The transformations $h_{i,j}$ and $h_{j,i}$ that register the images $T_i(x)$ and $T_j(x)$ are estimated by minimizing the cost function given by

$$\begin{aligned} C = \sigma \int_\Omega |T_i(h_{i,j}(x)) - T_j(x)|^2 + |T_j(h_{j,i}(x)) - T_i(x)|^2 dx \\ + \rho \int_\Omega ||\mathcal{L}u_{i,j}(x)||^2 + ||\mathcal{L}u_{j,i}(x)||^2 dx \\ + \chi \int_\Omega ||h_{i,j}(x) - h_{j,i}^{-1}(x)||^2 + ||h_{j,i}(x) - h_{i,j}^{-1}(x)||^2 dx \end{aligned} \quad (1)$$

where $u_{i,j}(x) = x - h_{i,j}(x)$ represents the displacement field from image $T_i(x)$ to $T_j(x)$. The first integral in the cost function defines the cumulative squared error similarity cost between the transformed template image $T_i(h_{i,j}(x))$ and the target image $T_j(x)$ and between the transformed target image $T_j(h_{j,i}(x))$ and the template image $T_i(x)$. This term indirectly defines the correspondences between the template and target images as the transformation that minimizes the intensity differences when each image is deformed into the shape of the other. The second integral in Eq. 1 is used to regularize the forward and reverse displacement fields $u_{i,j}$ and $u_{j,i}$, respectively. These terms constrain the transformations with a linear elasticity constraint. The linear differential operator $\mathcal{L}$ corresponding to the linear-elastic constraint given by $\mathcal{L}u(x) = \alpha\nabla^2 u(x) + \beta\nabla(\nabla \cdot u(x)) + \gamma u(x)$ where α, β, and γ are constants. The third integral in the cost function is called the inverse consistency constraint or the inverse consistency error. It is minimized when the forward $h_{i,j}$ and reverse $h_{j,i}$ transformations are inverses of each other. This term is important since it couples the forward and reverse transformations together forcing them towards the same correspondence mapping between the two images albeit inverses of one another. The parameters σ, ρ, and χ in Eq. 1 are weighting coefficients used to balance each term of the the cost function.

The top and bottom rows of Fig. 1 each show the results of a 2D image registrations obtained using the inverse-consistent linear-elastic image registration. Both registration results shown in this figure used the same template image (first column). The target image for both registrations are shown in the second column. The third and fourth columns show the template registered with the target image and the absolute intensity difference between the deformed template and target images, respectively. As is evident from the last column, these intensity differences are very small. Note that the target image was also deformed into the shape of the template but is not shown here for brevity of presentation.

Figure 2 shows the x and y-displacement fields and the log-Jacobian of the transformation from the template to the target images for the results shown in Fig. 1. The x and y-displacement fields show a circularly symmetric pattern for the registration of the average circle shape deformed into the image from the circle population. In contrast, the displacement fields show larger displacements near the location of the bump for the second registration result (bottom row). The log-Jacobian images show regions of contraction (white) and expansion (black) of the template image. Notice that the color scale for the top and bottom log-Jacobian images differ by an order of magnitude and that there are large regions of contraction and expansion in the area of the bump (bottom).

2.3 Synthesizing the Mean Shape of a Population

The T^2 statistic used to detect population shape differences is computed from the ICLEIR displacement fields at each voxel location in a reference coordinate system. The reference coordinate system used in this paper is defined as the average shape of the "normal" phantom population.

One image from the "normal" population without the bump was chosen as a template and registered with all the images in both the populations using inverse-consistent linear-elastic image registration (ICLEIR). The average shapes of both the populations were computed separately by transforming the template image with the average population

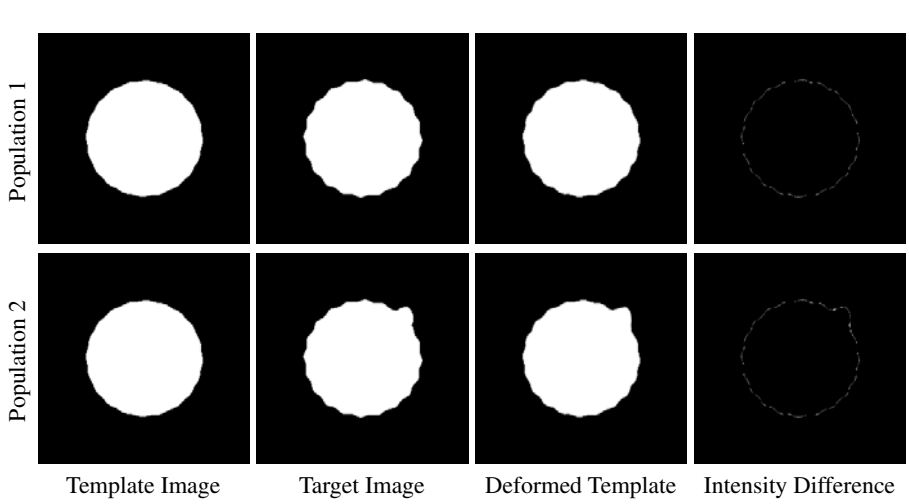


Fig. 1. Top and bottom rows show 2D inverse-consistent image registration results for an image from the population of circles without (top) and with (bottom) bumps. The first column shows the average image generated from the population of circles without bumps that was used as the template for both results. The second, third and fourth columns show the target image, the registered template image in the coordinate system of the target and the intensity difference between the target image and the registered template image, respectively. Note that each image is displayed on its own color scale.

displacement fields. Figure 3 shows the average shape obtained for the 30 data sets in each population. Notice that the average shape of the "normal" phantom population is very close in shape to a circle of radius 80. The right panel in Fig. 3 shows that the differences between the two computed average population shapes occurs at the location of the bump in the "abnormal" phantom population.

The average shape of a population of N images is computed using the method described in [8,9]. In this approach, an image i is selected from the population to be the template and is registered with all N images in the population using ICLEIR registration. This produces N Lagrangian transformations $g_{i,j}$ from image i to image j for $j = 1, \ldots, N$ as a function of the coordinate system of image i. The average Lagrangian transformation from the template image coordinate system to the population average coordinate system is defined as

$$g_{i,\bar{i}}(x) = \frac{1}{N} \sum_{j=1}^{N} g_{i,j}(x) \text{ for all } x \in \Omega. \tag{2}$$

The transformation $g_{i,\bar{i}}(x)$ defines a mapping from each coordinate location in the template coordinate system i to its corresponding location in the synthesized average coordinate system $\bar{i}$. To transform the the template image into the average coordinate system, the the Lagrangian transformation $g_{i,\bar{i}}(x)$ is converted into its Eulerian representation $h_{i,\bar{i}}(x)$ by computing the inverse of $g_{i,\bar{i}}(x)$, i.e., $h_{i,\bar{i}}(x) = \text{inv}\{g_{i,\bar{i}}(x)\}$. The population average image $\overline{T}_i$ is synthesized by deforming the template image T_i with the transformation $h_{i,\bar{i}}$ using the equation $\overline{T}_i(x) = T_i(h_{i,\bar{i}}(x))$. Linear interpolation is used to

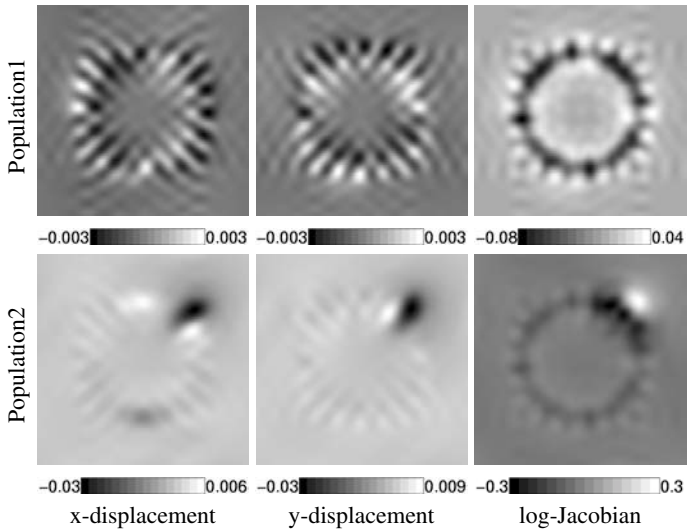


Fig. 2. Transformations and Log-Jacobian images for results shown in Fig. 1: from left to right: the x-displacement image, y-displacement image and the log-Jacobian of the transformation from the template coordinate system to the target image coordinate system.

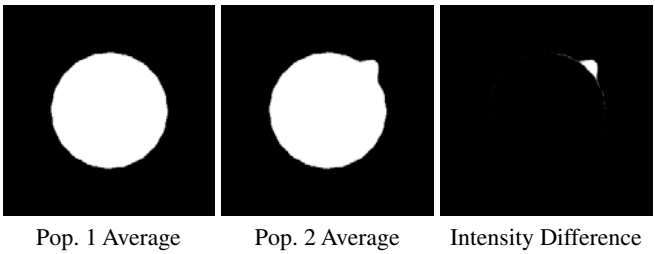


Fig. 3. The average images generated for the two populations and their intensity difference.

compute the intensity of each voxel of the image $\overline{T}_i(x)$ from the discrete template image $T_i(x)$.

2.4 Multivariate Hotelling's T^2 Hypothesis Testing

T^2 test was performed using the displacement fields as the parameters for comparing the shape differences. For our case of 2-dimensional images, m=2 and

$$X_{i,j} = (Xdisp_{ij}, Ydisp_{ij})^T \tag{3}$$

for each subject in each population. In the above equation, $Xdisp_{ij}$ and $Ydisp_{ij}$ represent the displacement in x-direction and displacement in y-direction in Lagrangian frame of reference from the template image i to subject j in the population. The covariance matrix S is a 2×2 square matrix.

The transformations from the "normal" average reference image to images in the "normal" and "abnormal" populations are used to perform the multivariate T^2 test. For populations 1 and 2 with n_1 and n_2 subjects respectively, the Hotelling's T^2 statistic based on m normally distributed variables (x_{11}, x_{12}, ... , x_{1n_1}) and $(x_{21}, x_{22}, ..., x_{2n_2})$ for the respective populations is calculated as

$$T^2 = (\bar{x}_1 - \bar{x}_2)^T \{S(1/n_1 + 1/n_2)\}^{-1} (\bar{x}_1 - \bar{x}_2) \tag{4}$$

where $X_{i,j} = (x_{ij1}, x_{ij2}, \ldots, x_{ijm})^T$ and $\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$. S is called the pooled covariance matrix and is defined as

$$S = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2} \tag{5}$$

where the population sample covariance matrices are defined as $S_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)^T$. The probability of being wrong in saying that populations 1 and 2 are different is given by

$$p = F^{-1}_{m, n_1+n_2-m-1}\left(\frac{n_1 + n_2 - m + 1}{m(n_1 + n_2 - 2)} T^2\right) \tag{6}$$

where F^{-1}_{m, n_1+n_2-m-1} is the inverse F-distribution with m and $n_1 + n_2 - m - 1$ degrees of freedom [1].

Very high values of the T^2 statistic corresponding to false positives occur when taking the inverse of an ill-conditioned S matrix. This condition occurs when there is little or no variation of the displacement fields across the populations such as in regions of constant intensity. In the present study, this problem occurred in the background and foreground of the reference image coordinate system.

This problem was addressed by adding a small constant $\epsilon > 0$ to the singular values of S before taking the inverse. The S matrix can be decomposed into its singular values using singular value decomposition given by

$$S = U \Lambda V^T \tag{7}$$

where U and V are orthogonal matrices and Λ is a diagonal matrix with non-negative diagonal elements. A new sample covariance matrix $\hat{S}$ with better conditioning is computed by adding a nonnegative constant ϵ to each diagonal element of Λ, i.e.,

$$\hat{S} = U(\Lambda + \epsilon I_m)V^T \tag{8}$$

where I_m is the $m \times m$ identity matrix. The sample covariance matrix $\hat{S}$ with improved conditioning is used to perform the statistical test. Conditioning the sample covariance matrix helps prevent false positives.

3 Results

Figure 4 shows regions of statistically significant regions of shape differences between the phantom populations for different ϵ and p-values. The results are shown as unthresholded and thresholded (1-p) images for different thresholds of 0.99 and 0.999. Note that these images are not binary.

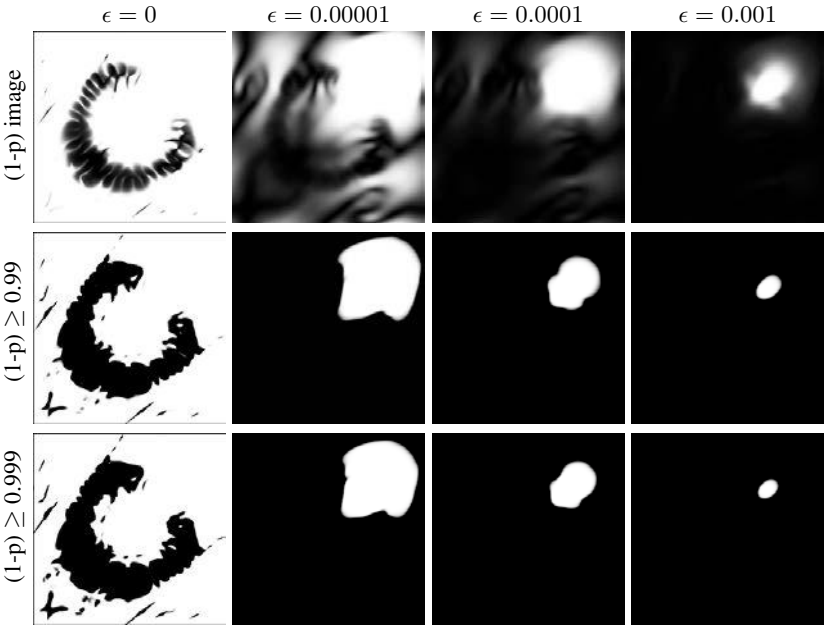


Fig. 4. Results of two population T^2 test for various values of ϵ and p-values.

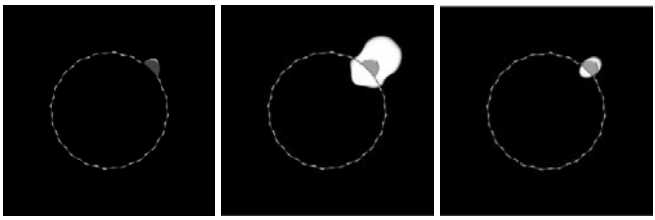
The first column corresponds to using the unconditioned S-matrix ($\epsilon = 0$). Notice that there are many false positives corresponding to the locations with very small variance in the displacement fields. Thresholding at (1-p) $\geq$ 0.99 and (1-p) $\geq$ 0.999 does not contribute much to separate these false responses from the regions of actual shape differences as shown in the second and third row for $\epsilon = 0$.

The second, third and fourth columns of Fig. 4 show that the false positive response region was reduced as ϵ increased. Notice that the selection of ϵ plays an important role in reducing the number of false positives. For a suitably high value of ϵ, the response of the statistical test can be tuned to the area of the desired response while minimizing the false positive responses. The response can be further focused by thresholding these images for different significance levels, i.e., (1-p) $\geq$ 0.99 and (1-p) $\geq$ 0.999. These results demonstrate that conditioning the S-matrix results in more meaningful population statistical shape comparisons.

Figure 5 overlays the response from the T^2 statistical test with the average intensity differences between the phantom populations. It is evident that as ϵ is increased from 0.00001 to 0.001, the response gets more focused to the region of statistically significant shape difference between the populations. Notice that for $\epsilon = 0.001$, the response region is focused to approximately the same size as the region with shape differences.

4 Summary and Conclusions

We presented a new method for detecting statistically significant shape differences between populations of images using a multivariate Hotelling's T^2 test on image corre-



A B C

Fig. 5. Panel A shows the intensity variance image for the pooled population. This figure was computed by calculating the variance of the intensity across all 60 population images at each pixel in the image. Panels B and C show the (1-p) ≥ 0.999 image superimposed on the pooled intensity variance image for $\epsilon = 0.0001$ and $\epsilon = 0.001$, respectively.

spondence maps. Inverse-consistent linear-elastic image registration (ICLEIR) [7] was used to determine individual to template correspondence maps and to compute the average shape of both populations [8,9]. Our approach improves the method proposed by Thirion et al. [6] by improving the conditioning of the sample covariance matrix for nearly singular sample covariance matrices used to compute the T^2 test.

We presented results comparing circle phantom populations of 30 images with bumps to a population of 30 images without bumps. This phantom study was designed to given us intuition as to how this test performs on data with known shape differences. We showed that improving the conditioning of the nearly singular sample covariance matrices dramatically reduced the number of false positives. It was shown that region of statistically significant shape differences was related to the conditioning number ϵ. This suggests that changing ϵ can be used to tune the method to detect statistically significant regions of shape difference. For the presented phantom experiment, it was shown that setting ϵ equal to 0.001 focused the response region to approximately the same size as the designed phantom shape difference.

This method is general and can be used to detect statistically significant shape differences between populations based solely on the correspondence maps. In the future, we plan on applying this method to real medical images and characterizing its performance on real data in a similar manner.

Acknowledgments

We would like to thank Russ Lenth, Director of the Statistical Consulting Center, The University of Iowa for his helpful comments. This work was supported in part by the NIH grants NS35368, DC03590, and HL64368. We would like to thank Richard A. Robb of the Mayo Clinic, Rochester, MN for his continued collaboration in providing the Analyze^{TM} software package.

References

1. R.A. Johnson and D.W. Wichern, *Applied multivariate statistical analysis*, Prentice-Hall, Upper Saddle River, N.J., 5 edition, 2002.

2. P.M. Thompson and A.W. Toga, "Detection, visualization and animation of abnormal anatomic structure with a deformable probabilistic brain atlas based on random vector field transformations," *Medical Image Analysis*, vol. 1, no. 4, pp. 271–294, 1997.
3. J. Ashburner and K.J. Friston, "Voxel-based morphometry - the methods," *NeuroImage*, vol. 11, no. 6, pp. 805–821, 2000.
4. I.C. Wright, Z.R. Ellison, T. Sharma, K.J. Friston, R.M. Murray, and P.K. McGuire, "Mapping of grey matter changes in schizophrenia," *Schizophrenia Research*, vol. 35, pp. 1–14, 1999.
5. I.C. Wright, P.K. McGuire, J.-B. Poline, J.M. Travere, R.M. Murray, C.D. Frith, R.S.J. Frackowiak, and K.J. Friston, "A voxel-based method for the statistical analysis of gray and white matter density applied to schizophrenia," *NeuroImage*, vol. 2, pp. 244–252, 1995.
6. Jean-Philippe Thirion and Neil Roberts, "Statistical analysis of normal and abnormal dissymmetry in volumetric medical images," *Medical Image Analysis*, vol. 4, pp. 111–121, 2000.
7. G.E. Christensen and H.J. Johnson, "Consistent image registration," *IEEE Transactions on Medical Imaging*, vol. 20, no. 7, pp. 568–582, July 2001.
8. G.E. Christensen and H.J. Johnson, "Synthesizing average 3D anatomical shapes," *Submitted. IEEE Transactions on Medical Imaging*.
9. G.E. Christensen, H.J. Johnson, J.W. Haller, M.W. Vannier, and J.L. Marsh, "Synthesizing average 3D anatomical shapes using deformable templates," in *Medical Imaging 1999: Image Processing*, K.M. Hanson, Ed., Feb. 1999, Proceedings of SPIE Vol. 3661, pp. 574–582.

A Gradient-Informed Robust Motion Correction Method for FMRI

Luis Freire[1,2,3] and Mark Jenkinson[1]

[1] Functional Magnetic Resonance Imaging of the Brain, Oxford University John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK
[2] Instituto de Biofísica e Engenharia Biomédica, FCUL, 1749-016 Lisboa, Portugal
[3] Instituto de Medicina Nuclear, HSM, 1649-028 Lisboa, Portugal
{lfreire,mark}@fmrib.ox.ac.uk
http://www.fmrib.ox.ac.uk

Abstract. Cerebral activation may be a serious confounding effect during the estimation of motion correction parameters in FMRI time series. This effect, which stems from the fact that activated voxels violate the assumption of intensity conservation for perfectly aligned images, is particularly significant when using least squares-based similarity measures. One way to deal with this problem is by down-weighting cerebral activation confounding signals during registration, which can be achieved using different metrics, other than least squares, based on robust estimators. However, this approach may lead to accuracy problems related to the increasing number of local minima, which manifest through an increased variability in motion estimates. The minimization of this problem could rely in the introduction of a pre-processing spatial smoothing step, but this strategy is likely to increase the bias between activation and motion correction parameters, due to the spatial consistency of the activation. A compromise between these two factors is obviously difficult.
In this paper, we present a different strategy, which consists of a gradient-informed robust motion correction method for FMRI time series. The robust similarity measure is least squares-based incorporating a Geman-McClure M-estimator. The cut-off power of the Geman-McClure estimator for each voxel pair is set as a linear function of the local gradient of the reference image. This strategy allows maintaining a high sensitivity relative to true intensity differences due to spatial misregistration, while minimizing activation-related confounding differences.
The robustness of the proposed method is first evaluated using a motion-free simulated time series including artificial activation-like signal changes based on a simple box-car paradigm. Results are compared with four other registration methods, which combine a least squares or a robust least squares similarity measure with a varying or non-varying strategy for spatial smoothing. These five methods are finally tested on three actual time series obtained from a 3T magnet.

1 Introduction

Patient motion may be a major source of error during the analysis of FMRI time series. The use of immobilization schemes, although useful to limit motion

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 377–386, 2003.

amplitude, is not sufficient to ensure an acceptable steadiness of patient's head. Therefore, neuroresearchers often rely on the retrospective registration of FMRI time series, in order to remove motion effects, which otherwise invalidate the later voxel-based statistical analysis.

The retrospective registration is normally performed by maximizing a similarity measure between each image that composes the time series and a chosen reference image (usually the first one after the scanner has reached the steady state). However, registration accuracy can be decisively affected by the presence of outliers due to cerebral activation. In a previous paper, it has been shown that activation presence may render motion compensation estimates falsely correlated with the activation paradigm [1]. The consequence is the appearance of spurious activations, mainly located near high-contrast edges.

One possible way to deal with such problems is through the use of robust metrics during the computation of the similarity measure. The aim of robust metrics is to try to minimize the influence of image dissimilarities (cerebral activation), assuming that near the optimal solution, their contribution to the similarity value is higher than the the contribution of the residuals due to misregistration. Therefore, one may attempt to decrease the influence of dissimilarities by reducing the contribution of larger residuals, hoping, at the same time, to maintain a sufficient number of true residuals in order to guarantee the accuracy of the method [2]. However, the accuracy of registration methods relying in robust metrics may be compromised due to the fact that the reduced number of voxels contributing to the similarity measure introduce additional problems to the optimization scheme (increased number of local minima), which lead to an increased variability in motion estimates.

In this paper, we propose a gradient-informed approach to the problem of robust metrics. The stringency of the robust estimator (in our case, a Geman-McClure (GM)) is not constant within the field-of-view, but it is a linear function of the local gradient magnitude of the reference image. This relation is inversely proportional, *i.e.* the higher the gradient, the less stringent the estimator. In order to exclude activation signals from gray matter voxels located at the brain surface, a stringent (constant) c-value for the estimator is used inside the brain, whilst immediately outside, where intensity changes are large due to the partial volume of bright brain with dark non-brain tissue, the gradient-informed (less stringent) strategy is used, so that these mismatches are allowed to drive the registration.

A second caution is related to the application of spatial smoothing in order to minimize local minima problems. This operation could easily introduce contaminations of voxels located outside the brain from activated voxels near brain edges. Therefore, our registration method follows a multiresolution prescheduled strategy for spatial smoothing, starting with strong smoothing and finishing with a minimum one. This strategy also helps to reduce the influence of activation bias, as demonstrated in [1]. This fact called for the evaluation of the performance of other registration methods, based on least square or robust

least square metrics, performing either the same multiresolution pre-scheduled strategy for spatial smoothing, or nore.

2 Materials and Methods

2.1 fMRI Acquisitions

In this paper, we have used three FMRI studies, obtained from three different subjects. The images were acquired on a Bruker scanner operating at 3T. Volume geometry consists of 18 contiguous slices (slice array 64 × 80), with in-plane resolution of 3.75 mm and slice thickness of 6.00 mm. The experimental paradigm was based on an on/off design consisting of two alternating visual stimuli, with period of 18 frames (2 s/frame). The time series include 10 periods plus 12 initial frames used to reach the scanner steady state.

2.2 Robust Metrics

The simplest metrics that can be used for image registration fall into the intensity conservation dependence taxonomy family defined in [2]. They correspond to the sum of absolute differences (SAD) and to the sum of squared differences (SSD). It is known that these metrics may not constitute the best choice when facing image dissimilarities that are not due to addictive random gaussian noise [3].

This limitation called for the development of robust metrics, specified through a ρ-function, which defines its properties in terms of cut-off power. The ρ-function, or estimator, may be seen as a transfer (or penalty) function, which down-weights the contribution of residuals above a certain level for the final similarity measure. The ρ-function depends on x, (x represents the residual value) but several other variables may be introduced, in order to describe different behaviors of the estimator (see [4]). The estimators shall have, ideally, moderate computation cost and a high breakdown point. The breakdown point is defined as the largest fraction of data that can be arbitrarily bad without causing the solution to be arbitrarily bad [5]. Among the different estimators proposed in the literature, the M-estimators deserve special interest. The M-estimators are defined by a symmetric ρ-function, monotonic in $\Re^+$. Robust metrics have been used in several works. Nikou uses M-estimators in least square-based and Woods similarity measures for MRI-MRI and MRI-SPECT registration [5]. Roche introduces an improvement of the CR similarity measure using several robust estimators [2].

The registration method we use is based on robust least square similarity measure that incorporates a Geman-McClure (GM) M-estimator. The GM estimator assumes the form $\rho(x) = x^2/(1+(x/c)^2)$; c is the variable that determines the cut-off power of the estimator.

2.3 Incorporating Gradient Information

The gradient information is used to assign different c values to different voxel locations. This strategy depends solely on the local gradient magnitude of the

voxel in the reference image, whatever the corresponding voxel in the test image. Therefore, we are interested in determining a gradient map to modify the c value (*i.e.* $c(r) = l(|g(r)|)$), in which $|g|$ is the local gradient magnitude, r the spatial coordinates, and l denotes a linear monotone increasing function. The c value ranges from 0.5% of mean brain value, for voxels in the reference image located in regions with low gradient, to 10% of the mean brain value (equivalent to SSD), for regions with high gradient.

The gradient map is obtained after applying a gaussian filter with 8 mm FWHM to the reference image. This option relies on the fact that one wishes to have a smooth gradient map, reducing the influence of noise.

As mentioned before, this gradient-informed strategy applies only to voxels located outside the brain, where the less stringent value for c near sharp edges allows eventually mismatches to drive the registration. For brain voxels, a non-informed strategy, based in a constant c value, is followed.

2.4 Registration Methods

As devised before, in this paper we test five different registration methods:

1. Least squares (LS1 and LS2): The least square-based registration methods are based in a SSD metric:

$$LS(A, B; T) = \sum_i (A_i - B_i^T)^2 \tag{1}$$

   in which B_i^T is the resampled value of image B at voxel i after the geometric transformation T has been applied.
   In LS1, the spatial smoothing step is performed using a 3-D Gaussian kernel with FWHM of 8 mm. In LS2, the FWHM is set to 8 mm during the first two stages, decreasing after that to 4 mm and finally to 1 mm. This multiresolution-like strategy for the spatial smoothing was implemented in order to minimize the contamination of voxels outside the head by neighboring activated voxels, and to reduce the activation-related bias in motion estimates.
2. Robust least squares incorporating a GM estimator (RLS1 and RLS2): The robust least squares method is calculated using the expression:

$$RLS(A, B; T) = \sum_i \frac{(A_i - B_i^T)^2}{1 + ((A_i - B_i^T)/c)^2} \tag{2}$$

   Both methods refer to a gradient non-informed strategy, *i.e.* the c value is constant for the entire field-of-view. As in the LS methods, RLS1 refers to a situation where the spatial smoothing is always the same (FWHM = 8 mm), and RLS2 to a multiresolution-like strategy for spatial smoothing, equivalent to LS2. The value of c was set the to 0.5% of mean brain intensity, according to previous experiments [6]. The small magnitude of the c value is due to the fact that activation-related confounding signals are typically in the range 1% to 4% of mean brain intensity.

3. Gradient-informed RLS (G-RLS): The incorporation of gradient information into the robust least square metric can be easily obtained from Equation 2, by setting $c' = l(|g|)$ for voxels located outside the brain and $c' = c$ for voxels inside the brain. Thus:

$$GRLS(A, B; T) = \sum_i \frac{(A_i - B_i^T)^2}{1 + ((A_i - B_i^T)/c')^2} \tag{3}$$

This method follows the same multiresolution-like strategy for spatial smoothing as the LS2 and RLS2 methods.

3 Experiments and Results

3.1 Simulated Time Series

The evaluation of the five registration methods is firstly performed using an artificial time series designed to simulate an activation process in the absence of subject motion. The experimental steps can be summarized as follows: (1) duplicate the reference image of the first actual time-series 40 times; (2) add an activation-like signal change in order to mimic a cognitive activation process. The activation time profile is shown in Figure 1; (3) Add Rician noise, obtained with two Gaussian distributions in real and imaginary axis with SD corresponding to 2% of mean brain value, in order to simulate the effects of thermal noise; (4) For each registration method, compute the Pearson's correlation coefficient between the 6 estimated motion parameters and the cognitive task profile. The added spatial activation pattern in (2) was obtained separately using SPM99, after resampling the first actual time series with motion estimates provided by a least square-based motion correction method. Therefore, the optimal registration method should yield a low correlation with the activation profile, combined with high accuracy (low error/variability) in motion estimates.

Table 1. Simulated time series: correlation values and mean RMS (obtained from the six registration parameters) for the five registration methods.

param.	LS 1	RLS 1	LS 2	RLS 2	G-RLS
t_x	0.38	0.35	0.26	0.16	0.05
t_y	0.93	0.60	0.53	0.54	0.03
t_z	0.92	0.33	0.40	0.05	0.08
r_x	0.93	0.46	0.65	0.11	0.14
r_y	0.07	0.19	0.21	0.43	0.15
r_z	0.38	0.25	0.23	0.12	0.15
RMS	0.013	0.009	0.006	0.025	0.010

From the results in Figure 2, one may see that the G-RLS method proposed in this paper can effectively reduce the systematic bias in motion estimates introduced by the presence of activation (a summary of the correlation coefficients

may be observed in Table 1). Also important is the fact that this is achieved mantaining a low variability in motion estimates (significantly smaller than the RLS2 method and similar to LS1 and RLS1). The root mean square (RMS) values, averaged for the six registration parameters, confirm this observation (0.010 for G-RLS and 0.009 for RLS1 against 0.025 for RLS2). These results are also presented in Table 1. The difficulties of the RLS2 method may be due to the fact that reducing smoothing factor combined with the decreasing number of voxels driving the similarity measure lead to optimization problems near the optimal solution.

On the other way, the LS2 method seems to perform better than the RLS2 method, despite some significant correlations, but with a very small amplitude. Besides, the RMS value for LS2 is the smallest (0.006), which indicates that the multiresolution-like strategy for spatial smoothing can, in this case, lead to a reduced bias influence with no additional problems for the optimization scheme. The LS1 method yielded the strongest correlations with the activation profile, which are clearly seen in t_y, t_z and r_x estimates.

3.2 Experiments with Real Data

The proposed method was applied to register the three real time series. For these data, the activation profile used to compute cross-correlation was obtained by convolving the task timing with the SPM99 hemodynamic model. A moving average was removed from the estimated motion parameters before computing the correlation in order to discard slow motion trend features, which are normally removed from the time series, prior to statistical analysis, by a high-pass preprocessing step.

The results obtained with the three actual time series also indicate a reduction in the correlation between the motion estimates and the activation paradigm, for the G-RLS method. See Table 2 for a summary of correlation coefficients. As can be seen in Figure 3, for the first real time series, the G-RLS also leads to a low variablity in motion estimates, specially when compared to RLS2 estimates. Indeed, one may see that the apparent robustness of RLS2 method against the presence of activation seems to result from the high variability of motion estimates, which suggests a lower accuracy of the RLS2 method. On the other way, the LS2 method yields a low variability in motion estimates, but a substantially increased correlation with the activation profile may be observed.

4 Discussion

The results presented in this paper support the idea that the proposed gradient-informed robust least square method provides a significant robustness against activation-related confounding signals during motion correction, with an accuracy level higher than the one that can be achived when using conventional (non-informed) robust metrics.

In the simulated experiments, one may observe that the multiresolution-like approach for the spatial smoothing used in LS2 method can effectively reduce

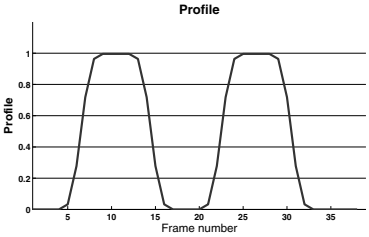


Fig. 1. Activation profile added to the simulated time series.

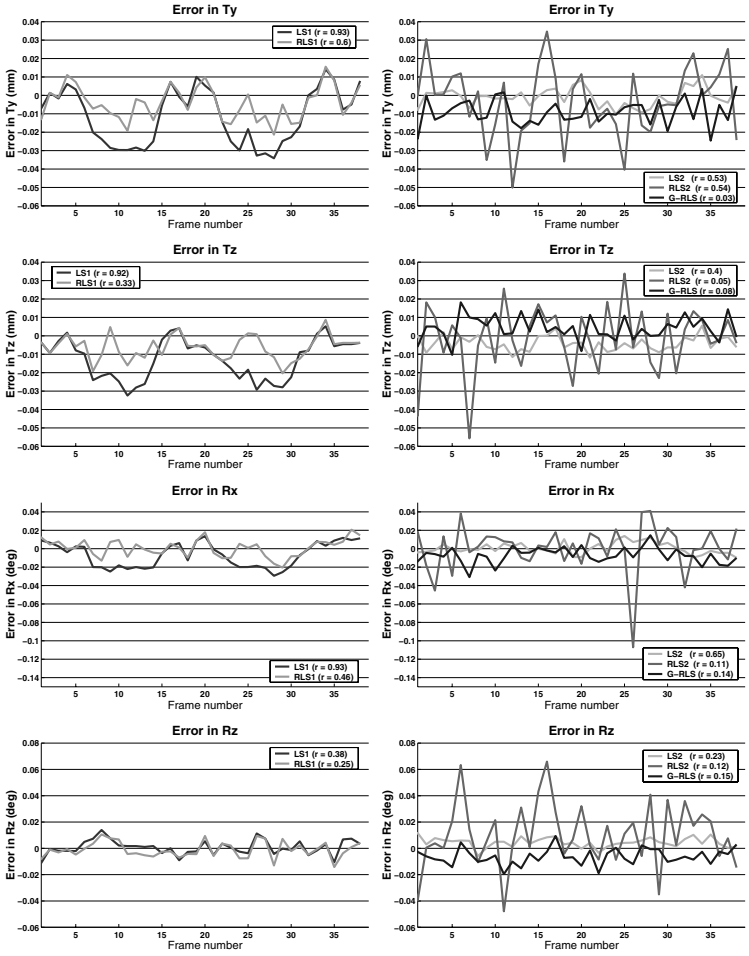


Fig. 2. Some of the registration parameters for the simulated time series. From *top* to *bottom*: t_y, t_z, r_x and r_z. Charts refer to LS1 and RLS1 (*left*) and to LS2, RLS2 and G-RLS methods (*right*).

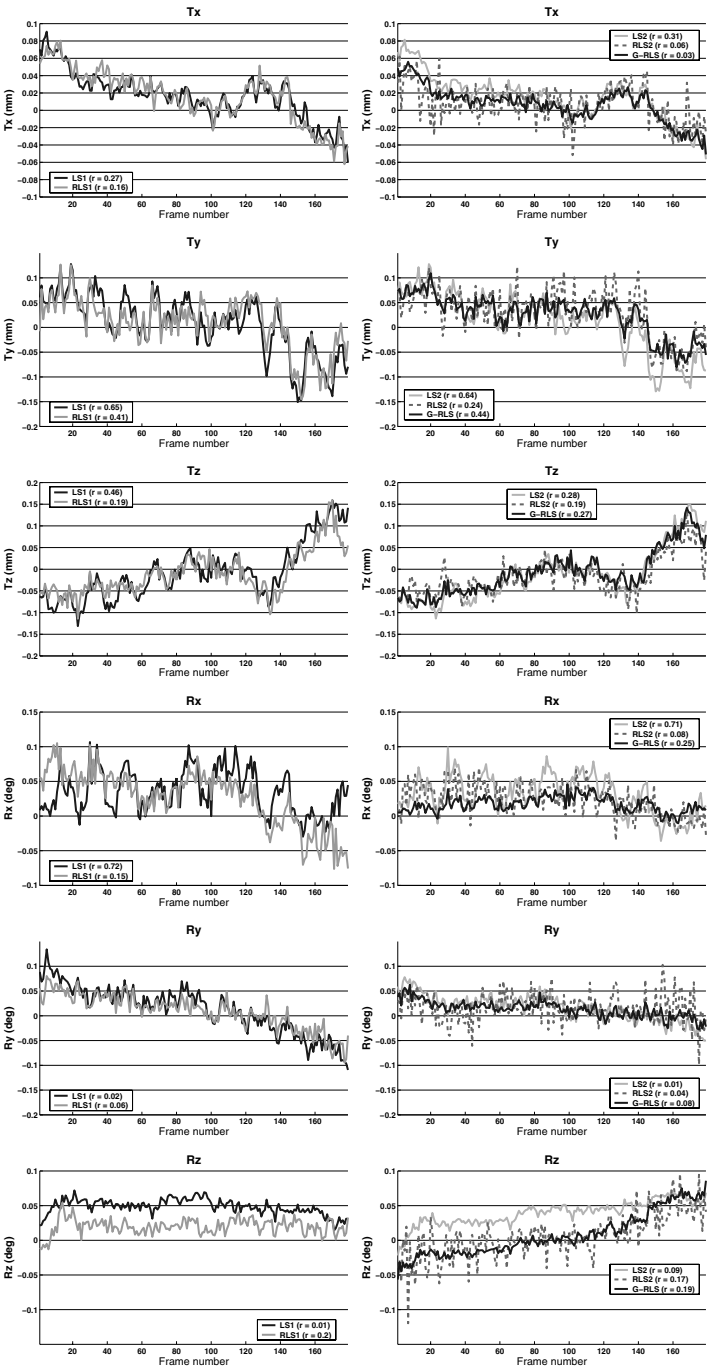


Fig. 3. Registration parameters for the first actual time series. Charts refer to LS1 and RLS1 (*left*) and to LS2, RLS2 and G-RLS methods (*right*).

Table 2. Correlation values for the three real time-series.

	param.	LS 1	RLS 1	LS 2	RLS 2	G-RLS
SET1	t_x	0.27	0.16	0.31	0.06	0.03
	t_y	0.65	0.41	0.64	0.24	0.44
	t_z	0.46	0.19	0.28	0.19	0.27
	r_x	0.72	0.15	0.71	0.08	0.25
	r_y	0.02	0.06	0.01	0.04	0.08
	r_z	0.01	0.20	0.09	0.17	0.19
SET2	t_x	0.17	0.05	0.12	0.01	0.15
	t_y	0.57	0.58	0.64	0.18	0.40
	t_z	0.63	0.24	0.47	0.05	0.10
	r_x	0.72	0.41	0.70	0.19	0.36
	r_y	0.05	0.15	0.17	0.12	0.21
	r_z	0.20	0.11	0.12	0.02	0.03
SET3	t_x	0.36	0.25	0.33	0.02	0.24
	t_y	0.67	0.55	0.53	0.11	0.26
	t_z	0.64	0.27	0.44	0.10	0.08
	r_x	0.69	0.29	0.64	0.10	0.29
	r_y	0.01	0.16	0.00	0.13	0.08
	r_z	0.38	0.17	0.38	0.02	0.04

the influence of outliers during the estimation of motion parameters, which is in accordance with the findings in [1]. Additionally, these results indicate that it provides a good protection against local minima problems. On the other way, we may see, from the results of RLS2 method, that this multiresloution-like approach for the spatial smoothing is not sufficient to eliminate the local minima problems associated with the robust least square method. Indeed, it seems that the with a constant c value, the RLS similarity measure introduces too many local minima problems for the optimization method. However, if the last smoothing stage is performed at 4 mm FWHM, the variability of motion estimates becomes substantially reduced but, at the same time, with an increase in the correlation coefficients with the experimental design (0.36, 0.66, 0.44, 0.42, 0.17 and 0.25 for t_x, t_t, t_z, r_x, r_y and r_z, respectively). In contrast, the G-RLS method yields very small correlation coefficients, while mantaining a low variability, close to the LS2 method.

Similar findings are obtained from the experiments with real time series. The substancial differences are related to the fact that the LS2 seems less robust against the presence of activation, which may be due to the low intensity of the activation-like signal in the simulated time series, and to the disagreemnet between motion estimates in the r_z estimate. This is probably due to spatial distortions induced by the fast acquisition scheme [7]. The consequence is the introduction of several local optima, which are due to local good matches between object features. A possible solution may rely in the use of an affine geometric model, capable to correct for geometric distortions. However, this choice is not free of compromises, since the increasing number of degrees of freedom may lead to additional local optima problems.

5 Conclusion

In this paper, we presented a new gradient-informed robust least square method for the problem of motion correction in FMRI time series. This method seems to provide a good robustness against the presence of activation, while maintaining a low motion estimate variability, which is often the main drawback of registration methods that down-weight outlier values. This work will be extended to evaluate the influence of the interaction between motion and susceptibility artifacts that occur in specific brain areas, which lead to signal loss and geometric deformations, for the final accuracy of the proposed gradient-informed robust least square method.

References

1. L. Freire and J.-F. Mangin, "Motion correction algorithms may create spurious brain activations in the absence of subject motion," *NeuroImage*, vol. 14(3) Sep, pp. 709–722, 2001.
2. A. Roche, *Recalage d'images médicales par inférence statistique, PhD Thesis.* Université de Nice-Sophia Antipolis, Projet Epidaure, INRIA, 2001.
3. P. Meer, D. Mintz, A. Rosenfeld, and D. Kim, "Robust regression methods for computer vision - a review," *International Journal of Computer Vision*, vol. 6(1) Apr, pp. 59–70, 1991.
4. M. Black and A. Rangarajan, "On the unification of line processes and outlier rejection and robust statistics with applications in early vision," *International Journal of Computer Vision*, vol. 19, pp. 57–91, 1996.
5. C. Nikou, F. Heitz, J.-P. Armspach, I.-J. Namer, and D. Grucker, "Registration of MR/MR and MR/SPECT brain images by fast stochastic optimization of robust voxel similarity measures," *NeuroImage*, vol. 8(1) Jul, pp. 30–43, 1998.
6. L. Freire and J.-F. Mangin, "Two-stage alignment of fMRI time series using the experiment profile to discard activation-related bias," in *5th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention, T. Dohi, R. Kikinis (eds.), Tokyo, Japan, LNCS 2489, Springer-Verlag, Berlin*, 2002, pp. 663–670.
7. C. Hutton, A. Bork, O. Josephs, R. Deichmann, J. Ashburner, and R. Turner, "Image distortion correction in fMRI: A quantitative evaluation," *NeuroImage*, vol. 16(1) Jul, pp. 217–240, 2002.

Clinical Applications from Head to Toe Using a Semiautomatic 3D Inter/Intramodality Fusion Technique

M.E. Noz[1], G.Q. Maguire Jr. [2], M.P. Zeleznik[3], E.L. Kramer[1], L. Olivecrona[4], H. Olivecrona[5], J. Crafoord[4], and J.K. Dewyngaert[6]

[1]Department of Radiology, New York University School of Medicine
New York, NY 10016, USA
{noz,kramer}@nucmed.med.nyu.edu; Keith.DeWyngaert@med.nyu.edu
[2]Institute for Microelectronics and Information Technology, Royal Institute of Technology
SE-100 44 Stockholm, Sweden
maguire@it.kth.se
[3]Department of Radiation Oncology, University of Utah, Salt Lake City, UT 84132 and RAHD Oncology Products, St. Louis, Mo. 63135, USA
zeleznik@cs.utah.edu
[4]Department of Radiology, Karolinska Hospital, SE-171 76 Stockholm Sweden
[5]Department of Orthopedics, Karolinska Hospital, SE-171 76 Stockholm Sweden
umh334t@tninet.se; joakim.crafoord@sverige.nu
[6]Department of Radiation Oncology, New York University School of Medicine
New York, NY 10016, USA

Abstract. Automated image registration techniques, particularly between modalities which clearly display anatomy and especially within the head, have become commonplace. Meanwhile advances in molecular imaging and the need to perform registration in other areas of body has driven the development of non-rigid and semiautomatic 3D volume fusion methods. This paper will focus on this latter class of 3D volume registration methods for a variety of clinical applications.

1 Introduction

Advances in medicine and patient care via diagnostic radiology, radiation treatment planning, surgery, and other methods of therapy, generally have benefitted from the data offered by using several different (generally complementary) imaging modalities. The choice of one modality over the other is governed by the information needed and the region of the body to be imaged. Typically volumetric modalities involving x-ray methods such as computer tomography (CT) and non-radiation methods such as nuclear magnetic resonance imaging (MRI) and ultrasound (US) provide good anatomic data. Volumetric molecular modalities, such as Magnetic Resonance Imaging Spectroscopy (MRIS), or those which involve the administration of radioactive pharmaceuticals (such as single photon emission tomography (SPECT) and positron emission tomography (PET)), are used to provide physiologic information.

The integration of information obtained from anatomic and/or physiologic data sets is often facilitated when some form of image fusion/registration[1,2] is used. Automated volume registration particularly between modalities which clearly display anatomy have become commonplace [3,4]. The situation becomes more complex when

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 387–397, 2003.

one of the anatomic modalities is replaced with a physiologic modality since the anatomic clues and other information required for successful, automated registration are generally not available, therefore operator intervention is often required [5-8]. Automated methods generally work very well in the brain[3,9], but have generally been less successful in the rest of the body[7]. Furthermore, automatic registration might not be desirable for all applications; e.g.,when the detection of prosthetic implant movement is desired [10,11].

In previous studies [12,13], the registration/fusion tool used here has been described and it has been shown that the 3D extensions to the original techniques[14,15] provide the ability to fuse all modalities *without* requiring specific acquisition techniques. This avoids the increased cost-per-use of fusion tools which require extensive time to acquire images (due to special positioning, fixtures, markers, …)[7]. Additionally, this tool is accurate and effective for all imaging modalities and applications, with few constraints on the original volumetric data [12] and hence difficulties encountered in registering images which are not initially well matched can be overcome. For example, the two volumes can be sliced at arbitrary angles, the patient can be oriented very differently in each volume (e.g., prone and supine), the modalities need not have strong data correlation (e.g., SPECT and CT), and external fiducials are not required. Furthermore, it has been shown that interactive methods can produce results comparable to automatic methods, [16] although there may exist specific cases which automatic methods can handle better - there is, as of yet, no automatic registration method which handles all cases better nor even always performs as well as the interactive method used here. The purpose of this report is to show the clinical advantages of using volume registration in the wide variety of applications to which we have applied this tool.

2 Methods and Materials

All of the studies described here were approved by the Institutional Review Board and informed consent was obtained when required. All studies were performed using *routine* clinical procedures and scanners.

All MRI scans were performed on one of three 1.5 T scanners (Vision or Symphony, Siemens Medical Systems, Iselin, NJ; Signa Advance, General Electric Medical Systems, Milwaukee WI). Gadolinium contrast material was used in some cases. The clinical protocols for all studies included both T1 and T2 weighted sequences which were reconstructed into 256 x 256 x 2 byte or 512 x 512 x 2 bytes matrices. The slice spacing range was 4.8 to 8.0 mm, and the x-y pixel size range was 0.78 to 0.98 mm.

All CT scans were acquired on one of three scanners (Highlight Advantage or High Speed, General Electric Medical Systems, Milwaukee WI, or PQ5000 Philips Medical Systems, Cleveland, Ohio). Oral and/or a bolus injection of nonionic contrast medium (Conray 43, Mallinkrodt, St. Louis, MO) followed by intravenous infusion, was administered. Between 30 and 50 transaxial sections were reconstructed into 512 x 512 x 2 byte matrices. The slice spacing range was 1.5 to 10.0 mm and the x-y pixel size range was 0.48 to 0.97 mm.

All PET scans were acquired on the same scanner (Advance PET, General Electric Medical Systems, Milwaukee, WI). Forty five minutes following the intravenous administration of 329 to 580 MBq (8.9 to 15.67 mCi) F-18 FDG tomographic studies

from the skull base through the proximal thigh were performed using a dedicated PET scanner. Emission scans were acquired for 5-6 minutes per bed position and transmission scans for 3 minutes per bed position. Between 190-210 slices were reconstructed into 128 x128 2 byte matrices, with and without attenuation correction. Only the attenuation corrected data sets were used. The slice spacing was 4.25 mm and the x-y pixel size was 4.30 mm.

The SPECT scans were acquired on one of three scanners (two- or three-headed gamma camera (BIAD, TRIAD, Trionix Research Laboratories Inc., Twinsborough, OH, or dual-headed gamma camera, GC7200, Toshiba America Medical Systems, Tustin, CA). Using the commercially supplied algorithm (filtered back projection with a Chang cylindrical attenuation correction and a Ramp post-filter), 90 to 120 projection slices were reconstructed into 60 to 70 slice volumes, contiguously covering the field of view. These data sets consisted of 128x128 x 2 byte matrices per slice of cubic voxels, which ranged from 2.0625 to 5.28 mm on each side.

The Coincidence Detection (CD) acquisition was performed with a dual headed gamma camera capable of coincidence detection (ADAC Laboratories, Milpitas, CA). Scans which included the head, lower head and neck, thorax, abdomen, and pelvis, were performed beginning one hour after the IV administration of approximately 185 MBq (5mCi) of F18-FDG in fasting patients. All sequences were reconstructed into a 128 x 128 x 2 byte matrices per slice of cubic voxels of 4.067 mm on a each side.

2.1 Head and Neck Studies

Data volumes were retrospectively gathered from six subjects (five women; one man; age range 21-71 years) with suspected recurrent papillary thyroid cancer. These patients had a history of papillary thyroid cancer and treatment with a near total thyroidectomy, and subsequent I-131 ablation of the remaining thyroid tissue. Although each patient had elevated thyroglobulin levels, no functioning tumor was seen on Iodine-131 whole body scans. F18-FDG CD and anatomic (5 MR; 1 CT) volume data sets were performed within 14 to 110 days of each other. These paired data sets were fused.

2.2 Thoracic Studies

Data volumes were retrospectively gathered from four patients (three women; one man; age range 58-77 years) with known or suspected lung carcinoma. Recurrent or new disease was suspected based on elevated normal markers. F18-FDG PET and CT volume data sets were fused. There was a range of 0 - 33 days between the two studies. Additionally, one woman with an unknown primary carcinoma underwent both an F-18 FDG Whole body PET and MRI scan of the breasts. The studies were performed within two days of each other. The MRI, limited to the breasts only and performed with the breasts pendant, was negative, but the PET, including the entire chest and with the patient supine, demonstrated uptake in a left axillary node.

2.3 Abdominal and Pelvic Studies

General Cancer Patients: Data volumes were retrospectively gathered from a group of 14 patients (12 women; two men; age range 7 - 69 years) who were being

evaluated for cancer, or for whom initial anatomic examinations had unexpected findings. Nine patients were given In-111 labeled B72.3, a tumor associated glycoprotein (TAG-72) specific-antibody, for the evaluation of ovarian cancer; one patient received In-111 MX-DTPA BrE-3, a mucin directed antibody for evaluation of advanced breast cancer; one patient was given In-111 octreoscan (Mallinckrodt), a somatostatin receptor ligand for detection of a suspected carcinoid, one patient with a history of a tumor of the left adrenal gland was studies with In-111 Octreoscan as well. Another patient received Tc-99m sestamibi for evaluation of a lung tumor; and, finally, another patient underwent Tc-99m sulfur colloid SPECT in the location of a suspected hepatocellular carcinoma. SPECT and anatomic (1 MR; 13 CT) volume data sets were fused [12].

Prostate Cancer Patients: At the time of these studies, there was a pool of 84 patients (age range 47 - 81years) who had undergone both In-111-radiolabeled capromab pendetide (ProstaScint, Cytogen Corp., Princeton, NJ), and anatomic (68 MRI; 16 CT) scanning. Most patients had undergone radical prostatectomy. They were referred for these studies because of a rising PSA. Capromab pendetide is a monoclonal antibody (MoAB) which adheres to prostate-specific membrane antigen (PSMA). All patients underwent planar and SPECT scans of the chest, abdomen and pelvis, 96 hours after infusion of 185-222 MBq (5-6 mCi) of In-111 labeled ProstaScint. On the day the ProstaScint scans were to be acquired, the patient was injected with 222 MBq (6 mCi) of Tc-99m red blood cells (RBC) so that a simultaneous blood pool scan could be acquired. The ProstaScint and RCB images were obtained concurrently. and, therefore, were spatially aligned.

Study one:
The impact of volume fusion on the interpretation of SPECT scans was evaluated. At the time, 68 patients (55 MRI; 13 CT) met the criterion of the study [17].

Study Two:
The purpose of this study was to demonstrate a method to extract meaningful biological information from ProstaScint SPECT scans for use in radiation therapy treatment planning by removing those components of the In-111 SPECT volumes associated with normal structure. For this study 20 patients (9 MRI; 11 CT) were selected from the above pool. All the patients in this group except one were post radical prostatectomy [18].

Patients with Hip Arthroplasties: Ten patients (six women; four men) that had undergone primary total hip arthroplasty (THA) with the Charnley implant system (DePuy, Warsaw, IN) were included in this study. A standardized protocol was designed to generate data on the entire pelvis and upper part of the femur without exposing the patients to a higher dose of radiation than routine postoperative plain radiography. Two CT scans were obtained from each patient in the first postoperative week. The scans where performed with an interval of 10 minutes. In between the scans the patient rose from the bed of the CT scanner and was subsequently repositioned. The acetabular component, i.e., the cup, of the prosthesis is made of polyethylene with a metal thread implanted to facilitate assessment of the cup position on radiographic examination. The metal thread is bent at approximately a 90° angle.

One hemicircle of the thread is designed to be parallel to the plane of the cup face and the other hemicircle to be parallel to the acetabular axis i.e. an axis that passes through the centre of the cup and is perpendicular to the plane of the cup face. The pelvis in each patient study was matched visually to within cortical width [10,11].

2.4 Evaluation Using Fixed Head-Frames in Patients Undergoing Stereotactic Surgery

Data volumes were gathered from 15 patients with brain tumors (six women; nine men; age range 17-71years) who had stereotactic frames attached to their skull. A stereotactic MRI scan and a Tl-201 SPECT scan were performed on the same day prior to the patient undergoing biopsy or surgery. In spite of the fact that the spatial coordinates of the two original scan volumes differed considerably, no preprocessing of the images was done before the actual matching. In the MRI scanner, each data set went from the top to the bottom of the brain, but did not include the top part of the frame. The stereotactic frame enclosing the patient's brain was attached to an apparatus that immobilized the patient's head within the field of view. For the SPECT scan, each patient was injected with 185 MBq (5 mCi) of Tl-201 chloride. A plastic frame insert was filled with Tc-99m and fitted to the stereotactic frame attached to the patient's skull. The front piece of the plastic frame was broken, and hence is not visualized. The SPECT volumes were acquired into two different energy channels, one containing only the frame (including the top and bottom of the frame) and the other, only the brain. In the SPECT scanner there was no apparatus to immobilize the patient's head [13].

2.5 Inter-patient MRI-MRI Abdominal Studies

Data volumes from three healthy, young men (age 25 years) were generated using MRI and matched in order to test if this tool could be used for inter-patient correlation. The long term goal is to generate and use a whole body atlas for radiation treatment planning [19].

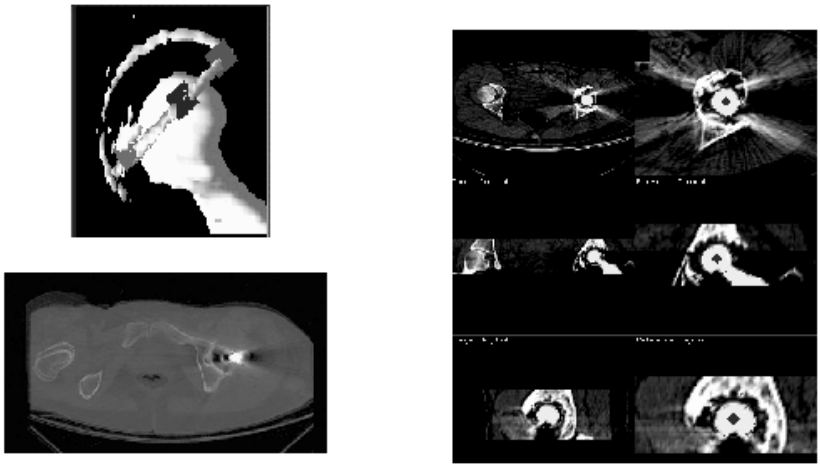
2.6 Data Transfer

The volume data sets were transferred via the network in DICOM 3 (CT, MRI, PET) or in original format (CD, SPECT) to a common computer system (HP 9000/C180V48 or HP C3000/fx6, Hewlett Packard Corp., Palo Alto, CA, or Sun UltraSPARC 10/Elite3Dm6, Sun Microsystems Inc., MountainView, CA) and converted to the standard AAPM format using Interformat [20]. Both volumes to be fused were read into the fusion program; either could be used as the reference or as the volume to be registered.

2.7 Volume Fusion

To register/merge any of these clinical studies, a previously described [10-13] volume fusion tool co-developed with RAHD Oncology Products was used. The registration algorithm incorporated in this tool can produce an affine or non-affine transformation. Arbitrarily chosen slices (with optionally superimposed isolines of the user's choos-

ing) are presented together in all three planes (axial, coronal, and sagittal) for both the anatomic and functional volume sets (six views) or as two larger views in one user-selected plane [12]. The slices can be individual panned and zoomed. Dual window width/level settings are optionally available. Each window can be displayed with a separate color scale. For CT, the lower window can be used for viewing the bony structures and soft tissue and the higher window can be used to simultaneously view metal or other high attenuating structures. An optional 3D shaded surface display, which can be arbitrarily panned, zoomed and rotated, is also available.

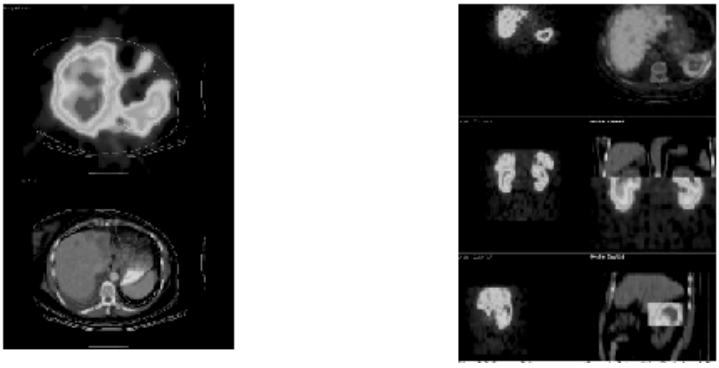


(A) (B)

Fig. 1. (A) Landmarks chosen on the prosthesis wire are checked in 3D (top); the matched pelvii are superimposed (bottom) - the patient's leg was positioned differently in each scan. (B) Illustration of using the spherical landmark to mark the center of the prosthetic femoral head. (Reprinted from [10] with permission, Blackwell Publishing Co.)

Co-homologous points, i.e., corresponding point pairs (landmarks), were chosen on concurrently viewed slices that display the same physiologic point or structure. Points could be selected in multiple planes simultaneously, using either a point or sphere if the exact center of the structure to be marked is desired. The sphere landmark superimposes the contours of a 3D sphere on the slices in all three planes (axial, coronal, and sagittal) together or separately. The size of the sphere can be chosen by explicitly specifying the radius before entering sphere mode or interactively with the mouse. When a landmark is chosen, the corresponding volume-element, recorded at the respective 3D position in distance units (mm), was marked, a sequential number generated, and all three planes triangulated to this point (Fig. 1).The eigenvalues of the matrix of coefficients were generated from the 3D paired landmarks by employing a weighted least squares linear regression followed by a Gauss-Jordan matrix inversion [13,14]. This procedure limits the effects of mismatched landmarks and provides an easy way to generate transformation coefficients for arbitrary volume data sets. Finally, polynomial equations formed from the eigenvalues, together with re-sampling,

determined the voxel values in the warped volume to bring it into correspondence with the reference volume, i.e., each original data point was moved into a transformed position. The transformation could be performed using only the first order terms (12 parameters in all) or using the second order terms (30 parameters in all) of the polynomial equations. Additionally an affine transformation requiring as few as four point pairs is provided. A completely manual (translate, rotate, and scale only) affine transformation was also available for either completely manual fusion or for use after the above transformation to bring the two volumes into better alignment (if required).



(A) (B)

Fig. 2. (A) SPECT slice with isolines (top) matched to the CT (bottom) with isolines from SPECT slice superimposed on CT. (B) Original SPECT slices (left) and superimposed on the CT (right) illustrating the various methods of visually evaluating the fusion. (B - reprinted from [12] with permission, Kluwer - Plenum Press.)

Evaluation of the fused volumes was done visually, numerically, and through clinical follow-up. For visual evaluation, the new transformed volume can be re-sliced and evaluated side by side with the reference volume in any of the three planes, or superimposed on (fused with) the reference slices (Fig 2). An isosurface of the transformed volume may be displayed in 3D, superimposed with an isosurface from the reference and/or the target volume for comparison. These 3D volumes can be rotated and viewed from an arbitrary direction, for visual evaluation. For numerical evaluation of the accuracy of the match, the original set of landmarks from each volumes to be registered were used. The same polynomial equations used to generate the warped volume were applied to these landmark sets. This produced a new set of landmarks that, if the transformation was perfect, would coincide exactly with the corresponding reference landmark positions. The distance between the actual position of the transformed landmarks and their ideal position (the reference landmark positions) was computed as well as the standard deviation of the sample, standard error of the mean, mean value, maximum value, and minimum value of the distance difference for each voxel (x, y, z), for each pixel (x, y), and for each direction separately. The volumes before and after registration were also cross correlated, the normalized correlation coefficient and the squared sum of voxel value differences were calculated. Additional numerical tests to assess overall registration accuracy have also been done [12, 11, 21].

3 Results

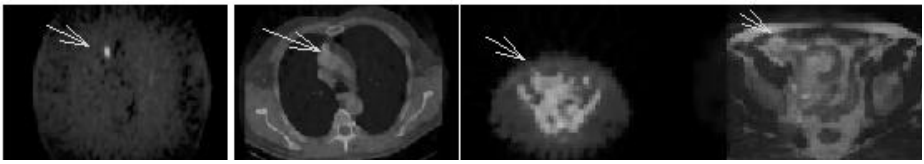
In all cases the fusion was accomplished within the clinically required parameters of match quality, time from start to finish of the process, ease of use, and level of training necessary to operate the software. A multi-variant ANOVA performed on the transformed landmark distance differences was not significant, confirming that the landmark choice between landmark sets and across patients was consistent. A t-test performed on these data showed in most cases that the transformed volume was within one to two pixels of the reference volume. The cross correlation, normalized correlation coefficient and squared sum of intensity differences always improved. In the brain-frame study we performed further statistical tests to show that the transformed volumes were significantly difference from the volumes before registration. For the orthopedic study, visual and numerical methods (as well as phantom studies [21]), confirmed that the method could locate the acetabular components within 1 mm. Therefore, a displacement of greater than 1 mm could be detected. Standard radiographic methods can only detect a displacement of greater than 4 mm. Clinical follow-up, where appropriate, confirmed the fusion findings (some positive, some negative) in 90% of the cases. Figs. 3-6 are presented to demonstrate the visual quality of the match.

4 Discussion

Volume fusion capability has advanced and matured considerably, especially in parts of the body other than the brain. We have demonstrated that functional and/or anatomic registration was achievable in the head, head and neck, thorax, abdomen, and pelvis. An added advantage of this tool is that it provides a 3D shaded surface display along with the 2D slices. By allowing the selection of landmarks in multiple planes, the user can view the physiologically/anatomically corresponding points from different points of view, which often strengthens the positive identification of corresponding features. Because the landmarks are saved and manipulated as 3D data points, there is no need for special acquisition protocols.

All studies were performed in a manner which makes them applicable to routine clinical use. The time for the complete process, reviewing data volumes, choosing landmarks, performing and accepting the transformation was 10 to 30 minutes. The user interface was rated by all the operators to be very easy to understand. Although the operator must have a good knowledge of anatomy, physicians (radiologists, orthopedic surgeon), residents, interns, medical students, and technologists have all repeatedly performed successful registrations.

It is very important when designing a volume registration system, to take into account such practical considerations, as these often determine the actual application domain and clinical applicability of the technique. A complete analysis evaluating the merits of a fusion tool should consider today's image acquisition technology and the changes that are certain to occur as technology advances; along with the financial implications of the real world. Many fusion tools limit the modalities that may be used, [5,6,9] or impose significant constraints for acquiring image data [7]. Obviously, such limitations of the fusion tool will decrease its routine clinical value as the clinician cannot easily incorporate new imaging techniques into their diagnostic or treatment planning process.

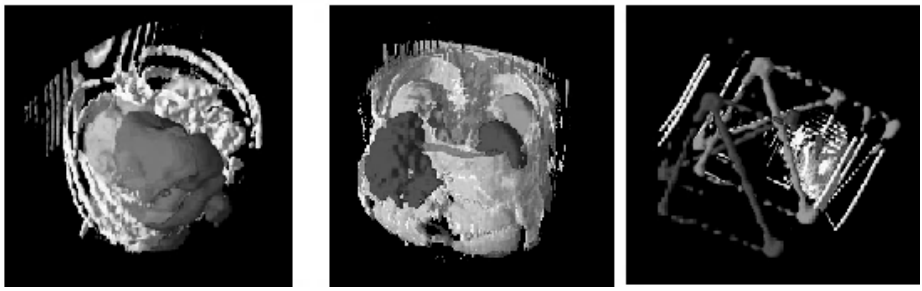


(A) (B)

Fig. 3. SPECT lesion on left (A) and (B); fused with anatomic view right (A) and (B)

(A) (B) (C)

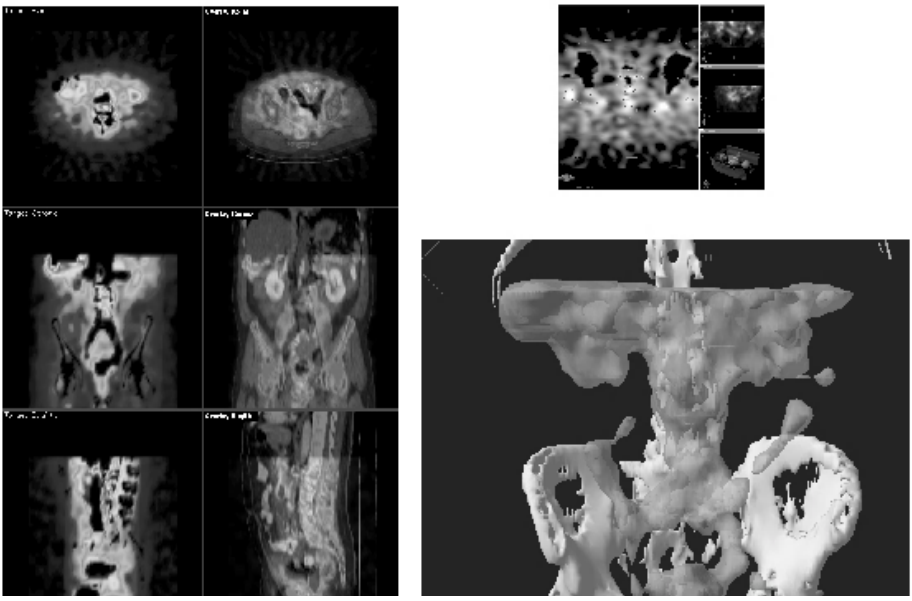
Fig. 4. CD head and neck volume merged with MRI showing anatomic location of the lesion; axial (A), coronal (B), and sagittal (C).



(A) (B) (C)

Fig. 5. 3D Match before and after warping, Original SPECT match dark grey, matched SPECT light grey, anatomic view white, (A) and (B) abdomen at the level of the liver/spleen (C) stereotactic frame match. (B reprinted from [12] with permission, Kluwer - Plenum Press).

Registration may not be needed for every patient undergoing functional and anatomic scans. It may only be required in those cases where there is a question regarding management and treatment, for example, confirmation of metastases versus fibrosis, accurate localization of active tumor tissue for biopsy or external beam radiation therapy planning, dosimetry for radioimmunotherapy, or precise prosthesis location. Our purpose in designing the 3D tool described above was to provide both sufficient speed and accuracy for such routine clinical uses.



(A) (B)

Fig. 6. (A) In-111 labeled ProstaScint SPECT with blood pool and bony structures removed (left) and fused with the CT (right). (B) SPECT volume presented using the treatment planning machine with contours superimposed from CT (top) and fused with the CT (bottom).

Acknowledgements

We thank C. Schettino, A. Murphy-Walcott, R.T.J.Gorniak, Y. Golowy, A. Kim, A. Deans, and F. Mahmoud for fusion and L. Wiedenheilm, L. Svensson, T. Jonson, for contributions to the orthopedic project.

References

1. Van den Elsen, P.A., Pol, E.J.D., and Viergever, M,A.: Medical Image Matching – a Review with Classification. IEEE Engineering in Medicine and Biology, EMB **40** (1993) 26-39,
2. Wahl, R., Quint, L., Cieslak, R., Aisen, A., Koeppe, R., Meyer, C: Anatometabolic tumor imaging: fusion of FDG PET with CT or MRI to localize foci of increased activity. Journal of Nuclear Medicine 34 (1993) 1190-1197
3. Maintz, J.B.A. and Viergever, M.A: A Survey of medical image registration. Medical Image Analysis **2** (1998) 1-36
4. Studholme, C., Hill, D.L.G., Hawkes, D.J.: Automated 3D Registration of MR and CT Images of the Head. Medical Image Analysis **1** (1996) 163-167
5. Studholme, C., Hill, D.L.G., and Hawkes, D.J.: Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures. Medical Physics **24** (1997) 25-35

6. Woods, R.P., Mazziotta, J.C., Cherry, S.R.: MRI-PET Registration with Automated Algorithm. Journal of Computer Assisted Tomography **17** (1993) 536-546
7. Klabbers, B.A., deMunck, J.C., Slotman, B.J.,Langendijk, H.A., deBree, R., Hoekstra, O.S.,Boellaard, R., and Lammertsma, A.A.: Matching PET and CT scans of the head and neck area: Development and method of validation. Medical Physics **29** (2002) 2230-2238
8. Pelizzari, C.A., Chen, G.T.Y., Spelbring, D.R., Weichselbaum, R.R., and Chen, C.T.: Accurate Three-Dimensional Registration of CT, PET and/or MR Images of the Brain. Journal of Computer Assisted Tomography **13** (1989) 20-26
9. Meyer, C.R., Boes, J.L, Kim, B., Bland, P., Zasadny, K.R., Kison, P.V., Kora, K., Frey,K.A, and Wahl, R.L.: Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin plate spline warped geometric deformations. Medical Image Analysis **3** (1996/7) 195-206
10. Olivecrona, L., Crafoord, J., Olivecrona, H., Noz, M.E., Maguire Jr., G.Q., Zeleznik, M.P., Svensson, L. Weidenhielm L.: Acetabular component migration in total hip arthroplasty using CT and a semi-automated program for volume merging. Acta Radiologica*,* **43** (2002) 517-527
11. Olivecrona, H., Weidenhielm, L., Olivecrona, L., Noz, M.E., Maguire Jr., G.Q., Zeleznik, M.P., Svensson, L., Jonson T.: Spatial component position in Total hip Replacement. Accuracy and repeatability with a new CT method. Acta Radiologica **44** (2003) 84-91
12. Noz, M.E., Maguire Jr., G.Q., Zeleznik, M.P., Kramer. E.L., Mahmoud F., Crafoord J.: A Versatile Functional-Anatomic Image Fusion Method for Volume Data Sets. Journal of Medical Systems 25 (2001) 297-307
13. Gorniak, R.J.T., Kramer, E.L., Maguire Jr., G.Q., Noz, M.E., Schettino, C.J., Zeleznik, M.P.: Evaluation of a semiautomatic 3D fusion technique applied to molecular imaging and MRI brain/frame volume data sets. Journal of Medical Systems **27** (2003) 141-156
14. Maguire Jr., G.Q., Noz, M.E., Rusinek, H., Jaeger, J., Kramer, E.L., Sanger, J.J., Smith, G.: Graphics Applied to Image Registration. IEEE Computer Graphics and Applications **11** (1991) 20-29
15. Katyal, S., Kramer, E.L., Noz, M.E., McCauley, D., Chachoua, A., and Steinfeld, A.: Fusion of Immunoscintigraphy Single Photon emission Computed tomography (SPECT) with CT of the Chest in Patients with Non-Small Cell Lung Cancer. Cancer Research (Suppl.) **55** (1995) 5759s-5763s
16. Pfluger, T., Vollmar, C., Wismüller, A., Dresel, S., Berger, P., Suntheim, P., Leinsinger, G., Hahn, K.: Quantitative Comparison of Automatic and Interactive Methods for MRI-SPECT Image Registration of the Brain Based on 3-Dimensional Calculation of Error. Journal of Nuclear Medicine **41** (2000) 1823-1829
17. Schettino, C.J., Noz, M.E., Kramer, E.L., Taneja, S., Lepor. H.: Impact of Image Fusion of the 111In Capromab Pendetide with MR or CT in Patients with Recurrent Prostate Ca (Abstr). Journal of Nuclear Medicine. **42** (2001) 294P
18. DeWyngaert, J.K., Noz, M.E., Ellerin, B., Kramer, E.L., and Walcott-Murphy, A.: Procedure for Unmasking Localization Information from ProstaScint Scans for Prostate Radiation Therapy Treatment Planning (Abstr*).* Int J Radiat Oncol Biol Phys 54 (2002) 32 (Suppl 1)
19. Crafoord, J., Mahmoud, F., Kramer, E.L., Maguire Jr.,G.Q., Noz, M.E. and Zeleznik, M.P.: Comparison of Two Landmark Based Image Registration Methods for Use with a Body Atlas. Physica Medica **XVI** (2000) 75-82
20. Reddy, D.P., Maguire Jr., G.Q., Noz, M.E., and Kenny, R.: Automating Image Format Conversion - Twelve Years and Twenty-five Formats Later. In: H.U. Lemke, K. Inamura, C.C. Jaffee, and R. Felix (ed). Computer Assisted Radiology - CAR'93. Berlin, West Germany: Springer-Verlag; (1993) 253-258
21. Olivecrona, L., Olivecrona, H., Weidenhielm, L., Noz, M.E., Maguire Jr., G.Q., Zeleznik, M. P.: Model studies on acetabular component migration in total hip arthroplasty using CT and a semi-automatic program for volume merging. Acta Radiologica **44** (2003) 11 pp.

A Parallel Implementation of Non-rigid Registration Using a Volumetric Biomechanical Model

M. Sermesant[1], O. Clatz[1], Z. Li[2], S. Lantéri[2], H. Delingette[1], and N. Ayache[1]

[1] Epidaure Research Project
[2] Caiman Research Project
INRIA Sophia-Antipolis, 2004 route des Lucioles
06902 Sophia-Antipolis, France

Abstract. Because of the accuracy required in a neurosurgical procedure, tracking intra-operative deformations is a challenging task. Furthermore, the clinical demand for fast non rigid registration will have to be met in a very near future. In this paper, we propose a patient-specific biomechanical model based on block-matching in order to register two MR images of the same patient with a parallel implementation. Compared to other intra-operative registration techniques, this method combines a viscoelastic mechanical regularization with a correlation-based iconic energy term. We first shortly present the theoretical aspects of our method (more detailed in [9]). Then we describe in more details the parallel implementation of the algorithm. Finally we present a retrospective registration study made of four pre/post operative MRI pairs.

1 Introduction

Non-rigid registration aims at deforming one image onto another image. However, even if the registered image has the same grey levels as the target image, the computed deformation may be very unrealistic. Capturing a realistic deformation field is important to follow critical structures.

In this paper, we introduce a biomechanical model in a registration framework, thus enabling to add anatomical and mechanical *a priori* knowledge. In this way, we indirectly constrain the displacement field to follow a realistic range of deformation.

Because biomechanical deformation computation can also be performed efficiently, it has been used to analyze per-operative brain deformations [2,10,7]. We propose a novel registration method which dynamically combines images and mechanical information. Our method relies on a dynamic law of motion of a tetrahedral mesh, which leads to iteratively solving a linear system of equations.

To have a computation time suitable for clinical use, we chose to rely on parallel computing since it has proved to be both powerful [11] and economically viable.

J.C. Gee et al. (Eds.): WBIR 2003, LNCS 2717, pp. 398–407, 2003.

2 Non-rigid Registration Method

The novelty in our registration algorithm is to integrate anatomical and mechanical *a priori* knowledge through a volumetric biomechanical model of the brain. The image information is used as force boundary conditions and external loads applied to the model. The aim of these forces is to deform the biomechanical model in such a way that the induced displacement field brings the reference image towards the target image.

In a deformable model framework, there is a balance between mechanical and iconic influences during the full registration process through an energy minimization.

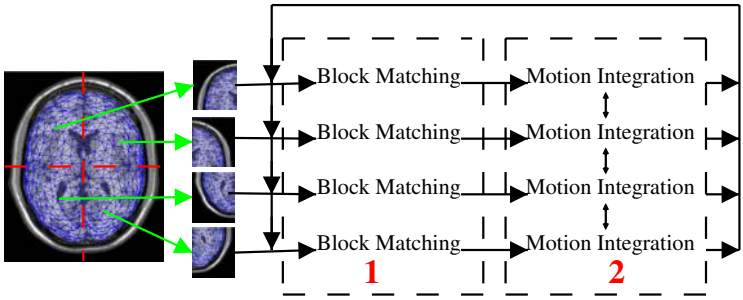


Fig. 1. Parallel implementation of the registration algorithm, distributing image backs and parallelizing the computation of image forces and the solution of a linear system of equations.

2.1 Internal Energy: Biomechanical Model

The internal energy regularizes the displacement field and is based on continuum mechanics, more precisely on linear elasticity. This simple model guarantees that an area will follow the motion of its neighbors without image matching computed in this area.

2.2 External Energy: Block-Matching Displacements

The external energy is based on intensity correlation such that each vertex is moved by an image force towards a voxel of similar intensity in the target image. Until now, deformable models mainly used gradient or distance map-based forces applied on surface vertices along the normal. We propose to use a volumetric external energy based on 3D block matching for two reasons :

- full motion of the organ can be registered, not limited to apparent motion, as with forces along the surface normal.
- more complex deformation can be estimated inside the model, not limited to interpolation of the surface deformation.

The quality of the registration with such a volumetric external energy does not rely entirely on the surface segmentation accuracy, unlike most surface-based methods.

The external energy is based on the assumption that the intensity around a given vertex of the mesh should be similar in the two images, the difference only coming from noise and anatomical changes. Thus we attach to each vertex a surrounding block of voxels from the reference image and search in the target image for the most similar block around the current position. The matching voxel selected is the center of the block maximizing the chosen similarity criterion. Then, an external force computed for each vertex is proportional to the distance to the matching voxel, and directed from the vertex current position towards the position of that voxel.

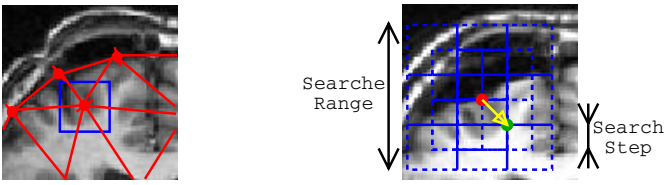


Fig. 2. (Left) Reference Image: 3D blocks of voxels initialized for each vertex in the reference image. (Right) Target Image: surrounding blocks searched for similarity measure maximum.

Several similarity measures can be used (sum of square differences, mutual information,...) depending on the type of noise and image modalities used. Moreover, this type of external energy is well-suited for a coarse-to-fine approach, by varying the search step and the search range in the target image [5].

3 Motion Integration

It is important to take into account the inertia effect, and therefore to use a dynamic equation of motion. Indeed, even if the update of the displacement field is small between two time-steps, it is much more efficient and stable to find a balance between external and internal energies with a dynamic approach.

Therefore, we integrate the motion in the dynamic equation:

$$M\frac{d^2U}{dt^2} + C\frac{dU}{dt} + KU = F \tag{1}$$

with M, C, K, the mass density ($= 1$ for the brain), damping, stiffness matrices (taken from [1]) and U the displacement from rest position ($P(t) = P(0) + U(t)$, with $P(t)$ the position at time t). F is the external force vector computed from the image matching.

3.1 Numerical Integration

We use the finite element method to numerically solve this system. Spatially, we use linear tetrahedral elements and temporally, the semi-implicit Houbolt method. With this method, speed and acceleration are discretized as follow:

$$\frac{d^2P(t+\Delta t)}{dt^2} = \frac{1}{\Delta t^2}[2P(t+\Delta t) - 5P(t) + 4P(t-\Delta t) - P(t-2\Delta t)]$$

$$\frac{dP(t+\Delta t)}{dt} = \frac{1}{6\Delta t}[11P(t+\Delta t) - 18P(t) + 9P(t-\Delta t) - 2P(t-2\Delta t)]$$

Replacing this in eq. (1) gives a linear system to solve in $U(t+\Delta t)$ at each iteration. Since we are looking for small deformations ($< 10\%$) and small displacements [3], the stiffness matrix is chosen constant (linear elasticity).

The registration process can be decomposed in two steps:

1. the initialization procedure:
   - mass M, damping C and stiffness K matrices assembly,
   - $U(0)$, $\dot{U}(0)$, $\ddot{U}(0)$ initialization,
   - time step choice,
   - several constants computation:

$$a_0 = \frac{2}{\Delta t^2}\,;\; a_1 = \frac{11}{6\Delta t}\,;\; a_2 = \frac{5}{\Delta t^2}\,;\; a_3 = \frac{3}{\Delta t}\,;$$
$$a_4 = -2a_0\,;\; a_5 = \frac{-a_3}{2}\,;\; a_6 = \frac{a_0}{2}\,;\; a_7 = \frac{a_3}{9}$$

   - $U(\Delta t)$ and $U(2\Delta t)$ initialization with $U(0)$,
   - effective stiffness matrix computation:

$$\hat{K} = K + a_0 M + a_1 C$$

   - $\hat{K}$ preconditioner construction. Since the stiffness matrix is constant (because the time step is constant), this operation is performed only once.

   As this step should be computed pre-operatively, the corresponding execution time has not been taken into account in the results section. However, this full step takes about 3 seconds on a 1 Ghz Pentium III PC.
2. the solution of a linear system of equations at each time step:
   - effective external forces computation:

$$\hat{F}(t+\Delta t) = F(t+\Delta t) + M\,(a_2U(t) + a_4U(t-\Delta t) + a_6U(t-2\Delta t))$$
$$+ C(a_3U(t) + a_5U(t-\Delta t) + a_7U(t-2\Delta t))$$

   - iterative linear system solution:

$$\hat{K}U(t+\Delta t) = \hat{F}(t+\Delta t) \tag{2}$$

   - nodes positions update.

4 Implementation

4.1 Mesh Creation

The different steps of the automatic process used to create a tetrahedral mesh of the brain from a MRI are:

1. rigid registration to align the mid-sagital plane of the MRI with a coordinate plane, if the image is not already registered along a coordinate, using the algorithm proposed in [6],
2. cortex binary mask creation, using the method proposed by Mangin [4] and implemented in Brainvisa[1]. This software, downloadable on the web, is fully automatic and very efficient,
3. creation of a brain triangle mesh, from the mask isosurface extraction. The number of triangles chosen for this surface is directly related to the number of tetrahedra in the final mesh
4. creation of a brain tetrahedral mesh, using the software GHS3D[2], developed at INRIA by the Gamma Research Project.

This fully automated mesh creation step takes less than one minute (for a mesh with ≈ 40000 tetrahedra) on a 1 Ghz Pentium III PC.

4.2 Mesh Partition

The parallelization of the registration process relies on a distributed memory model. Each processor only computes the biomechanical and image forces for a subpart of the mesh. The tetrahedral mesh is partitioned using the METIS[3] software, a state-of-the-art graph partitioning tool that also includes fill-reducing reordering capabilities. At the end of the partitioning step, each vertex of the mesh has been assigned to a processor in such a way to minimize the communication cost.

4.3 Matrix Assembly

The PETSc[4] library is used for parallel matrix and vector storage, preconditioning and iterative system solving. The stiffness matrix is computed, preconditioned and distributed at the beginning of the process. For message passing operations, PETSc relies on the widely adopted Message Passing Interface (MPI)[5]. Moreover, PETSc enables to work only with global mesh vertex indices even though, internally, local indices are used for accessing the distributed data structures.

[1] http://www.brainvisa.info/
[2] http://www-rocq.inria.fr/gamma/ghs3d/ghs.html
[3] http://www-users.cs.umn.edu/~karypis/metis/
[4] http://www-unix.mcs.anl.gov/petsc/petsc-2/
[5] http://www.erc.msstate.edu/misc/mpi/index.html

4.4 Linear System Preconditioning

Different preconditioning methods are available in PETSc. Because image-based forces produce a volumetric field that can be very discontinuous, the force vector can be quite different form typical physics-based forces. Therefore, the most efficient preconditioners can be different from those used in classical linear elasticity problems. Moreover, even if the standard incomplete Cholesky factorization method probably offers the best compromise in terms of computational and memory costs, a more efficient and more costly preconditioner can be used here since the preconditioning step is performed only once for the overall registration process.

4.5 Linear System Solving

A 3D problem in not tractable for a direct system solver: the inverse of a sparse matrix is not sparse, and the required computational and memory costs to obtain and store the inverse can be prohibitive. In this context, preconditioned iterative solvers are very efficient to find an approximation of the solution.

Many iterative methods are available. Even if the stiffness matrix is symmetric positive definite, it appeared that solvers non-specific to such matrices can be more efficient as illustrated in table 1 below.

Table 1. Preconditioners and solvers comparison for a mesh of 2000 nodes on a PC equipped with an Intel Pentium IV/2 GHz processor. CG: Conjugate Gradient, GMRES: Generalized Minimum RESidual, BICGSTAB: BIConjugate Gradient STABilized, IC: Incomplete Cholesky, ILUT: Incomplete Lower Upper with Threshold.

Solver	*Preconditioner (time)*	*Iterations Number*	*Total solution Time*
CG	IC (0.04 s)	19	0.414 s
GMRES	ILUT (1.63 s)	11	0.387 s
BICGSTAB	ILUT (1.63 s)	7	0.365 s

Each processor computes the external forces for the vertices corresponding to the associated submesh. Then, the distributed right hand side of system (2) is computed, and the system is iteratively solved.

5 Results

5.1 Performance Evaluation

We propose to evaluate the performance of the proposed algorithm on synthesized images. As proposed in [8], we use the finite element method to generate a gravity-induced deformation on a virtual brain from BrainWeb[6]. The average displacement of the induced deformation is 11,6mm and the maximum is 27mm.

[6] `http://www.bic.mni.mcgill.ca/brainweb/`

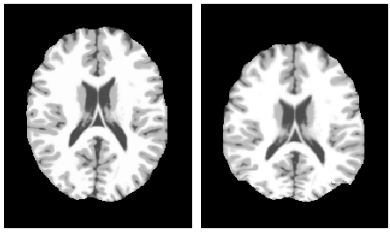


Fig. 3. Gravity-induced deformation with the FEM method.

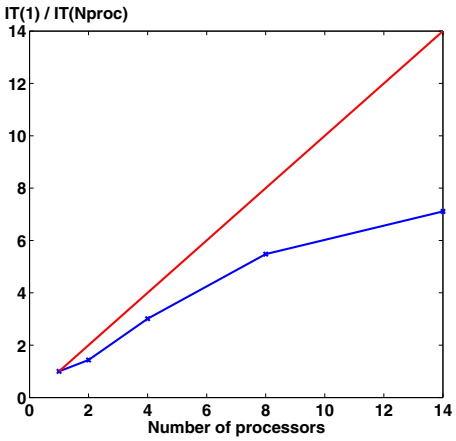


Fig. 4. Red: Ideal, Blue: f(N CPU) = Iteration Time (1 CPU) / Iteration Time (N CPU), Range of block-matching research = 5mm, research step = 2mm.

This can be considered as the largest possible deformation during a real neurosurgical procedure (see figure 3). We finally add noise to both deformed and non-deformed images and test our algorithm on this final pair of images.

Figure 4 shows the influence of the number of processors on the execution time (denoted by IT on the figure) of one outer iteration step (i.e. time step). These peformance results have been obtained on a cluster of PCs consisting of 14 Intel Pentium IV/2 Ghz processors interconnected by a Gigabit-Ethernet switch. The underlying tetrahedral mesh of the brain contains around 2000 vertices. The straight line on this figure defines the ideal parallel speedup. The rapid decrease of the parallel efficiency is simply due to the relatively coarse mesh used in our test.

We then studied the influence of the damping factor on the convergence of the algorithm. We can see on figure 5 that the dynamic effects of the damping factor is a critical point in the registration computation process. Indeed, a small damping factor leads to an instability whereas a large damping factor leads to a large computation time. We manually estimate the right damping factor to 0,015, whereas we do not have automatic way to determine it. The optimum damping factor may actually depend on the searched deformation. This damping factor

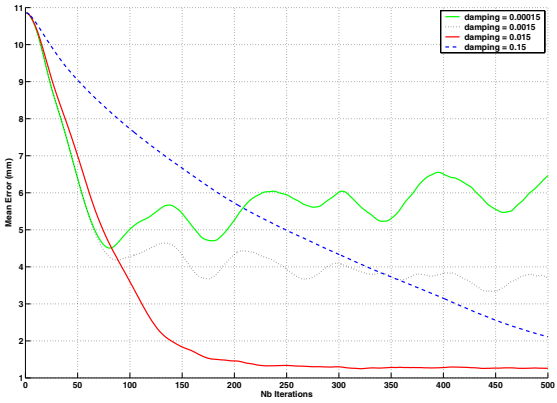


Fig. 5. Influence of the damping factor on the convergence and the mean error (in millimeters). Without enough damping, the convergence is not obtained, due to the instability. With too much damping, the convergence is importantly slowed.

Table 2. Influence of the noise on the robustness of the algorithm (white Gaussian noise on intensity, variable standard deviation sigma).

Sigma	0	1	2	3	4	5	6	7
Average final error (mm)	1,31	1,33	1,33	1,33	1,35	1,33	1,43	1,41

authorizes an average error of 1,22mm (10, 5% error) in less than 3 minutes on 14 CPU.

We present in table 2 a study of the robustness to noise of the algorithm. We can see that the algorithm, presented here with the correlation coefficient similarity measure, performs well with a high degree of noise on images. We can even consider it to be nearly not affected by noise in the common range of noise on an MRI ($3 < \sigma < 5$)

5.2 Tests on Real Images

The following results have been obtained with the fully automatic registration process previously described. In addition to the brain creation stage, we removed all the image-based forces applied on the brain surface. This implies that the surface layer of tetrahedra only move with the biomechanical regularization. In that way, the registration result is independent of the segmentation quality. Each image have been computed in less than a minute (block-matching searching range = 5mm, searching step = 2mm, correlation coefficient, 30 000 tetrahedra in the mesh, 14 CPUs).

The results obtained on real images are very encouraging. The volumetric biomechanical registration seems to be well suited to recover a physical deformation. The difference image shows the good ability of the algorithm to deform internal structures. In addition, the algorithm performs well on the brain surface,

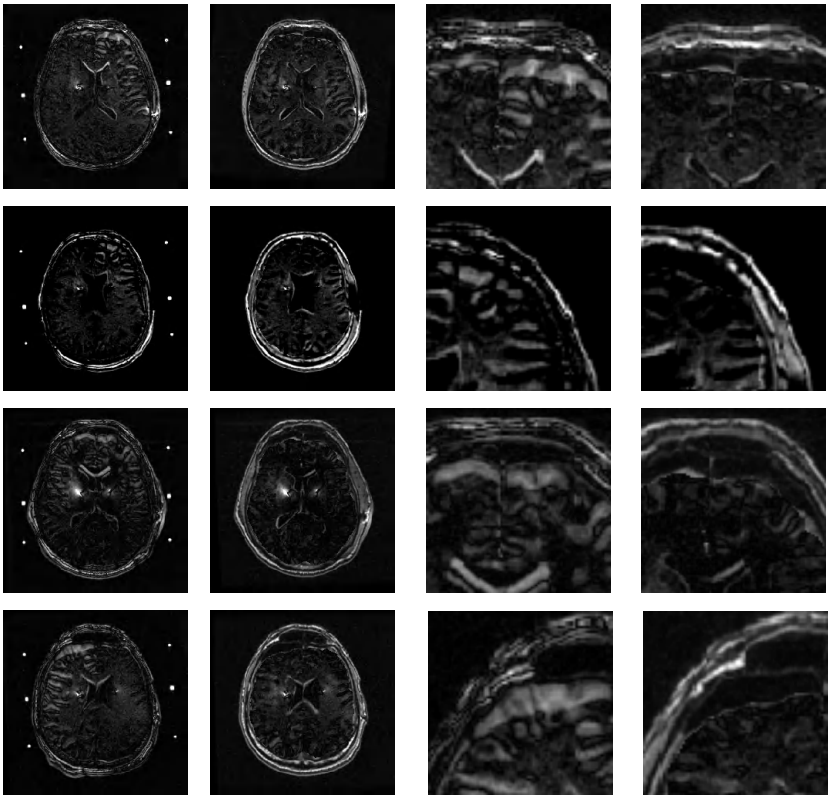


Fig. 6. Difference between pre- and post-operative images. 1st column: before registration. 2nd column: after registration. Zoom on the largest displacement area. 3rd column: before registration. 4th column: after registration. Images come from the Neuroradiology Department, La Pitié Salpêtrière Hospital, Paris, courtesy of Pr. D. Dormont.

even with no image criterion. This ability to recover the surface displacement with only internal constraints shows the efficiency of the volumetric external energy.

The artifacts due to the electrodes produce registration errors in the surrounding areas. However we can be removed from their influence in the block matching search. In this way we obtain a completely smooth displacement field in this area.

6 Conclusion

Using a biomechanically constrained registration method leads to a more robust method for two reasons:

- a better regularization in outliers areas thanks to the biomechanical model smoothing
- a more realistic deformation field thanks to the anatomical and mechanical knowledge included

However these advantages are computationally costly. A parallel implementation makes an intra-operative use of such methods conceivable.

Ongoing work on a dedicated PC cluster should make precise time, precision and robustness comparisons possible. Moreover, mechanical parameters, like damping, can then be extensively tested and adjusted. And on a dedicated PC cluster, the communication time is very reduced, as the network is not slowed down by other tasks.

We also wish to pursue the work on synthesized deformations by studying the influence of the size of the mesh on the global error.

References

1. O. Clatz, H. Delingette, E. Bardinet, D.Dormont, and N. Ayache. Patient-specific biomechanical model of the brain: application to Parkinson's disease procedure. In *Surgery Simulation & Soft Tissue Modeling (IS4TM'03)*, 2003.
2. M. Ferrant, A. Nabavi, B. Macq, P. Black, F. Jolesz, R. Kikinis, and S. Warfield. Serial registration of intraoperative mr images of the brain. *Medical Image Analysis*, 6(4):337–360, 2002.
3. S. Kyriacou, A. Mohamed, K. Miller, and S. Neff. Brain mechanics for neurosurgery: Modeling issues. *Biomechanics and Modeling in Mechanobiology*, 2002.
4. J.-F. Mangin, O. Coulon, and V. Frouin. Robust brain segmentation using histogram scale-space analysis and mathematical morphology. *Lecture Notes in Computer Science*, 1496:1230–1241, 1998.
5. S. Ourselin, A. Roche, S. Prima, and N. Ayache. Block Matching: A General Framework to Improve Robustness of Rigid Registration of Medical Images. In A.M. DiGioia and S. Delp, editors, *Third International Conference on Medical Robotics, Imaging And Computer Assisted Surgery (MICCAI 2000)*, volume 1935 of *Lectures Notes in Computer Science*, pages 557–566, Pittsburgh, Pennsylvanie USA, octobre 11-14 2000. Springer.
6. S. Prima, S. Ourselin, and N. Ayache. Computation of the mid-sagittal plane in 3D brain images. *IEEE Transaction on Medical Imaging*, 21(2):122–138, February 2002.
7. J. Rexilius, S. Warfield, C. Guttmann, X. Wei, R. Benson, L. Wolfson, M. Shenton, H. Handels, and R. Kikinis. A novel nonrigid registration algorithm and applications. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI'01)*, volume 2208 of *LNCS*, pages 923–931. Springer, 2001.
8. J. A. Schnabel, C. Tanner, A. Castellano-Smith, M. O. Leach, C. Hayes, A. Degenhard, R Hose, D. L. G. Hill, and D. J. Hawkes. Validation of Non-Rigid Registration using Finite Element Methods. In *In Proc. Information Processing in Medical Imaging (IPMI'01)*, volume 2082 of *Lectures Notes in Computer Science*, pages 344–357, University of California at Davis, 18-22 June 2001. Springer.
9. M. Sermesant, O. Clatz, H. Delingette, and N. Ayache. Non-rigid registration combining volumetric biomechanical model and block-matching. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI'03)*, Lecture Notes in Computer Science (LNCS). Springer, 2003. submitted.
10. O. Skrinjar, A. Nabavi, and J. Duncan. Model-driven brain shift compensation. *Medical Image Analysis*, 6(4):361–374, 2002.
11. S. Warfield, F. Talos, A. Tei, A. Bharatha, A. Nabavi, M. Ferrant, and R. Kikinis P. Black, F. Jolesz. Real-time registration of volumetric brain MRI by biomechanical simulation of deformation during image guided neurosurgery. *Computing and Visualization in Science*, 5(1):3–11, 2002.

Author Index