Bjoern Menze Georg Langs
Zhuowen Tu Antonio Criminisi (Eds.)

Medical Computer Vision

Recognition Techniques and Applications in Medical Imaging

LNCS 6533

International MICCAI Workshop, MCV 2010
Beijing, China, September 2010
Revised Selected Papers

Springer

Lecture Notes in Computer Science 6533

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Bjoern Menze Georg Langs Zhuowen Tu
Antonio Criminisi (Eds.)

Medical Computer Vision

Recognition Techniques and Applications in Medical Imaging

International MICCAI Workshop, MCV 2010
Beijing, China, September 20, 2010
Revised Selected Papers

Springer

Volume Editors

Bjoern Menze
CSAIL - Computer Science and Artificial Intelligence Laboratory, MIT
Cambridge, MA 02139, USA
E-mail: menze@csail.mit.edu

Georg Langs
CSAIL - Computer Science and Artificial Intelligence Laboratory, MIT
Cambridge, MA 02139, USA
E-mail: langs@csail.mit.edu

Zhuowen Tu
University of California, Laboratory of Neuroimaging
Los Angeles, CA 90095-7334, USA
E-mail: zhuowen.tu@loni.ucla.edu

Antonio Criminisi
Microsoft Research, Cambridge CB3 0FB, United Kingdom
E-mail: antcrim@microsoft.com

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-18420-8 e-ISBN 978-3-642-18421-5
DOI 10.1007/978-3-642-18421-5
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2010942789

CR Subject Classification (1998): I.4, I.2.10, H.3, I.5, J.3

LNCS Sublibrary: SL 6 – Image Processing, Computer Vision, Pattern Recognition, and Graphics

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The Workshop on Medical Computer Vision (MICCAI-MCV 2010) was held in conjunction with the 13th International Conference on Medical Image Computing and Computer – Assisted Intervention (MICCAI 2010) on September 20, 2010 in Beijing, China. The one-day workshop focused on recognition techniques and applications in medical imaging. The participants discussed principled approaches that go beyond the limits of current model-driven image analysis, which are provably efficient and scalable, and which generalize well to previously unseen images. It included emerging applications that go beyond the analysis of individual clinical studies and specific diagnostic tasks – a current focus of many computational methods in medical imaging.

The workshop fostered discussions among researchers working on novel computational approaches at the interface of computer vision, machine learning, and medical image analysis. It targeted an emerging community interested in pushing the boundaries of what current medical software applications can deliver in both clinical and research medical settings. Our call for papers resulted in 38 submissions of up to 12 pages each. Each paper received at least three reviews. Based on these peer reviews, we selected 10 submissions for oral and 11 for poster presentation.

The *Best Scientific Paper Award* was given to Michael Kelm and co-authors for their contribution "Detection of 3D Spinal Geometry Using Iterated Marginal Space Learning." The runners-up were Peter Maday and co-authors for their contribution "Imaging as a Surrogate for the Early Prediction and Assessment of Treatment Response Through the Analysis of 4-D Texture Ensembles," and Dongfeng Han and co-authors for their contribution "Motion Artifact Reduction in 4D Helical CT: Graph-Based Structure Alignment."

During the workshop two distinguished invited speakers offered their perspective on the development of the field: Dorin Comanicu of Siemens Corporate Research, and Yiqiang Zhan of Siemens Medical Solutions.

September 2010

Bjoern Menze
Georg Langs
Zhuowen Tu
Antonio Criminisi

Organization

MCV 2010 was organized during the MICCAI 2010 Conference at Beijing.

Workshop Chairs

Workshop Chairs	Bjoern Menze (MIT, INRIA) Georg Langs (MIT) Zhuowen Tu (UCLA) Antonio Criminisi (Microsoft Research)

Program Committee

Padmanabhan Anandan	Microsoft Research
Nicolas Ayache	INRIA
Christian Barillot	IRISA
Horst Bischof	TU Graz
Katja Buehler	VRVIS
Dorin Comaniciu	Siemens
Tim Cootes	University of Manchester
Rachid Deriche	INRIA
Polina Golland	MIT
Hayat Greenspan	University of Tel Aviv
James Hays	Brown University
Joachim Hornegger	University of Erlangen
Michael Kelm	Siemens
Ron Kikinis	Harvard
Ender Konukoglu	Microsoft Research
Koen Van Leemput	Harvard
Hans-Peter Meinzer	DKFZ Heidelberg
Albert Montillo	Microsoft Research
Théodore Papadopoulos	INRIA
Nikos Paragios	ECP, INRIA
Xavier Pennec	INRIA
Killian Pohl	IBM
Tammy Riklin-Raviv	MIT
Karl Rohr	DKFZ Heidelberg
Robert Sablatnig	TU Vienna
Yonggang Shi	UCLA
Philipp Torr	Oxford University
Martin Urschler	TU Graz
Tom Verkauteren	Mauna Kea Technologies

René Vidal	Johns Hopkins
Michael Wels	Siemens
Xiang Sean Zhou	Siemens

Sponsoring Institutions

Microsoft Research

Table of Contents

Shape, Geometry and Registration

Conditional Point Distribution Models 1
Kersten Petersen, Mads Nielsen, and Sami S. Brandt

Deformable Registration of Organic Shapes via Surface Intrinsic Integrals: Application to Outer Ear Surfaces 11
Sajjad Baloch, Alexander Zouhar, and Tong Fang

Iterative Training of Discriminative Models for the Generalized Hough Transform 21
Heike Ruppertshofen, Cristian Lorenz, Sarah Schmidt, Peter Beyerlein, Zein Salah, Georg Rose, and Hauke Schramm

Topology Noise Removal for Curve and Surface Evolution 31
Chao Chen and Daniel Freedman

Exploring Cortical Folding Pattern Variability Using Local Image Features 43
Rishi Rajalingham, Matthew Toews, D. Louis Collins, and Tal Arbel

Markov Models for Image Reconstruction and Analysis

Surgical Phases Detection from Microscope Videos by Combining SVM and HMM 54
Florent Lalys, Laurent Riffaud, Xavier Morandi, and Pierre Jannin

Motion Artifact Reduction in 4D Helical CT: Graph-Based Structure Alignment 63
Dongfeng Han, John Bayouth, Sudershan Bhatia, Milan Sonka, and Xiaodong Wu

Comparative Validation of Graphical Models for Learning Tumor Segmentations from Noisy Manual Annotations 74
Frederik O. Kaster, Bjoern H. Menze, Marc-André Weber, and Fred A. Hamprecht

Automatic Anatomy Localization via Classification

Localization of 3D Anatomical Structures Using Random Forests and Discrete Optimization 86
René Donner, Erich Birngruber, Helmut Steiner, Horst Bischof, and Georg Langs

Detection of 3D Spinal Geometry Using Iterated Marginal Space Learning 96
B. Michael Kelm, S. Kevin Zhou, Michael Suehling, Yefeng Zheng, Michael Wels, and Dorin Comaniciu

Regression Forests for Efficient Anatomy Detection and Localization in CT Studies 106
Antonio Criminisi, Jamie Shotton, Duncan Robertson, and Ender Konukoglu

Correcting Misalignment of Automatic 3D Detection by Classification: Ileo-Cecal Valve False Positive Reduction in CT Colonography 118
Le Lu, Matthias Wolf, Jinbo Bi, and Marcos Salganicoff

Learning Adaptive and Sparse Representations of Medical Images 130
Alessandra Staglianò, Gabriele Chiusano, Curzio Basso, and Matteo Santoro

Feature Selection for SVM-Based Vascular Anomaly Detection 141
Maria A. Zuluaga, Edgar J.F. Delgado Leyton, Marcela Hernández Hoyos, and Maciej Orkisz

Texture Analysis

Multiple Classifier Systems in Texton-Based Approach for the Classification of CT Images of Lung 153
Mehrdad J. Gangeh, Lauge Sørensen, Saher B. Shaker, Mohamed S. Kamel, and Marleen de Bruijne

Imaging as a Surrogate for the Early Prediction and Assessment of Treatment Response through the Analysis of 4-D Texture Ensembles (ISEPARATE) 164
Peter Maday, Parmeshwar Khurd, Lance Ladic, Mitchell Schnall, Mark Rosen, Christos Davatzikos, and Ali Kamen

A Texture Manifold for Curve-Based Morphometry of the Cerebral Cortex 174
Maxime Boucher, Alan Evans, and Kaleem Siddiqi

Semisupervised Probabilistic Clustering of Brain MR Images Including Prior Clinical Information 184
Annemie Ribbens, Frederik Maes, Dirk Vandermeulen, and Paul Suetens

Segmentation

Simultaneous Multi-object Segmentation Using Local Robust Statistics and Contour Interaction 195
Yi Gao, Allen Tannenbaum, and Ron Kikinis

Spotlight: Automated Confidence-Based User Guidance for Increasing Efficiency in Interactive 3D Image Segmentation 204
Andrew Top, Ghassan Hamarneh, and Rafeef Abugharbieh

Automated Segmentation of 3D CT Images Based on Statistical Atlas and Graph Cuts 214
Akinobu Shimizu, Keita Nakagomi, Takuya Narihira, Hidefumi Kobatake, Shigeru Nawano, Kenji Shinozaki, Koich Ishizu, and Kaori Togashi

Author Index 225

Conditional Point Distribution Models

Kersten Petersen[1], Mads Nielsen[1,2], and Sami S. Brandt[2]

[1] Department of Computer Science, University of Copenhagen, Denmark
[2] Synarc Imaging Technologies, Denmark

Abstract. In this paper, we propose an efficient method for drawing shape samples using a point distribution model (PDM) that is conditioned on given points. This technique is suited for sample-based segmentation methods that rely on a PDM, *e.g.* [6], [2] and [3]. It enables these algorithms to effectively constrain the solution space by considering a small number of user inputs – often one or two landmarks are sufficient. The algorithm is easy to implement, highly efficient and usually converges in less than 10 iterations. We demonstrate how conditional PDMs based on a single user-specified vertebra landmark significantly improve the aorta and vertebrae segmentation on standard lateral radiographs. This is an important step towards a fast and cheap quantification of calcifications on X-ray radiographs for the prognosis and diagnosis of cardiovascular disease (CVD) and mortality.

1 Introduction

Many segmentation problems in medical image analysis are challenging, because the image is afflicted by clutter, occlusion or a low signal-to-noise ratio. Several techniques have been proposed that exploit local appearance and global shape information to account for these difficulties. However, a fully automated approach reaches a limit, when there is too much uncertainty in the correct solution, or when the segmentation problem is ill-posed. In these cases, a human expert can assist by refining the input given to the segmentation algorithm. For instance, in vertebrae segmentation it frequently occurs that the vertebrae boundary is accurately found, but displaced by one vertebral level [7]. Here, a single known point on the vertebrae boundary could resolve the ambiguity and clearly improve the overall segmentation performance.

In this work, we propose an algorithm for drawing shape samples using a point distribution model (PDM) [4] that is conditioned on fixed landmark points. The technique can be beneficial for segmentation methods that are based on geometric sampling from a PDM, *e.g.* [6] and [3]. Our goal is that a human expert guides these methods to the correct solution by fixing a few landmark points. Usually one or two points should be sufficient to constrain the solution space effectively. We will demonstrate the accuracy and effectiveness of our algorithm on an important medical application: the vertebrae and aorta segmentation on lateral X-ray radiographs for the prognosis and diagnosis of cardiovascular disease and mortality.

B. Menze et al. (Eds.): MICCAI 2010 Workshop MCV, LNCS 6533, pp. 1–10, 2011.

The remainder of this paper is organized as follows. We recapitulate the idea of (unconditioned) PDMs, as described in [4], before explaining the idea of conditional PDMs in Section 2,. Section 3 presents our experimental results to evaluate our method. In Section 4 we will discuss the potential of conditional PDMs and conclude.

1.1 Point Distribution Model

A point distribution model (PDM) is a statistical shape model which can be built from a collection of training shapes. Each input shape of the PDM is given by a set of landmarks points which are in correspondence across all the training shape instances. In the training phase, we first align the landmarks to a mean shape with the generalized procrustes algorithm [5]. Then we perform principal component analysis (PCA) to compute the modes of shape variation. Usually we keep only a small number of principal components, so that most of the shape variation is retained, while the risk of overfitting is lowered. Hence, the PDM represents each shape by four pose and a few shape parameters: the pose parameters describe the similarity transformation from measurement to shape space, whereas the shape parameters determine the projection on the chosen principal components.

2 Problem Definition

Consider a vector $\mathbf{x} = (x_1, \ldots, x_N, y_1, \ldots, y_N)^{\mathrm{T}} \in \mathcal{X}$ describing the N landmark points of a two-dimensional shape in measurement space $\mathcal{X} = \mathbb{R}^{2N}$. Let us further assume that $\mathbf{x}$ can be divided into a known part $\mathbf{x}_K$ and an unknown part $\mathbf{x}_{\setminus K}$, where $K = (k_1, \ldots, k_M, k_1 + N, \ldots, k_M + N)^{\mathrm{T}}, \{k_m\}_{m=1}^{M} \in \{1, \ldots, N\}$ indexes the coordinates of M known landmark points and $\setminus K$ denotes the complementing indices.

Our objective is to draw shape samples from a conditional PDM $p(\mathbf{x}_{\setminus K}|\mathbf{x}_K)$ that is related to the distribution of shapes $\tilde{\mathbf{x}}$ in the shape space $\tilde{\mathcal{X}} = \mathbb{R}^{2N}$. The similarity transformation $T : \mathbb{R}^{2N} \times \mathbb{R}^4 \to \mathbb{R}^{2N}$ parameterized by $\theta = (s, \alpha, t_x, t_y)$ maps the coordinates $\tilde{\mathbf{x}}_K$ to the fixed coordinates $\mathbf{x}_K$, where s is the scale, α the rotation angle, and (t_x, t_y) the 2D translation. By writing shape vector $\mathbf{x}$ as a $2 \times N$ matrix $\mathbf{X}$ such that each row represents a shape coordinate, $\mathbf{R}$ the rotation matrix and $\mathbf{T}$ the translation matrix, the similarity transformation takes the form

$$\mathbf{X} = s\mathbf{R}\tilde{\mathbf{X}} + \mathbf{T} \,, \tag{1}$$

Introducing $\mathrm{vec}\{\mathbf{X}\}$ to indicate the vectorization of matrix $\mathbf{X}$ along the rows yields

$$T(\tilde{\mathbf{x}}, \theta) \equiv \mathbf{x} = \mathrm{vec}\{\mathbf{X}\} = \mathrm{vec}\{s\mathbf{R}\tilde{\mathbf{X}} + \mathbf{T}\} \,. \tag{2}$$

Figure 1 illustrates the similarity transformation from shape to measurement space. Note that the shown aorta samples are represented as a set of landmarks in the dual space, the two-dimensional image space.

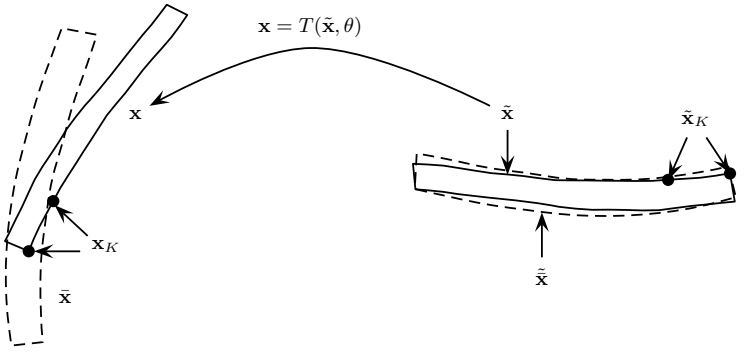


Fig. 1. The similarity transformation $T(\tilde{\mathbf{x}}, \theta)$ maps an (aorta) sample $\tilde{\mathbf{x}}$ from shape space $\tilde{\mathcal{X}}$ to measurement space $\mathcal{X}$. The dashed aorta denotes the mean shape in the respective space. Note that this image shows the samples as a set of landmarks in the dual space, the two-dimensional image space.

Let us define the combined pose and shape vector $\mathbf{z} = (\tilde{\mathbf{x}}, \theta) \in \mathcal{Z}$, where $\mathcal{Z}$ denotes the shape space, and use $\mathbf{\Sigma_z}$ to denote the associated covariance matrix from the PDM. Then sampling from an unconditioned PDM corresponds to sampling from a normal distribution

$$p(\mathbf{z}) \propto \exp\left(-\frac{1}{2} f(\mathbf{z})\right) \tag{3}$$

$$f(\mathbf{z}) \equiv (\mathbf{z} - \bar{\mathbf{z}})^{\mathrm{T}} \mathbf{\Sigma}_{\mathbf{z}}^{-1} (\mathbf{z} - \bar{\mathbf{z}}) \,, \tag{4}$$

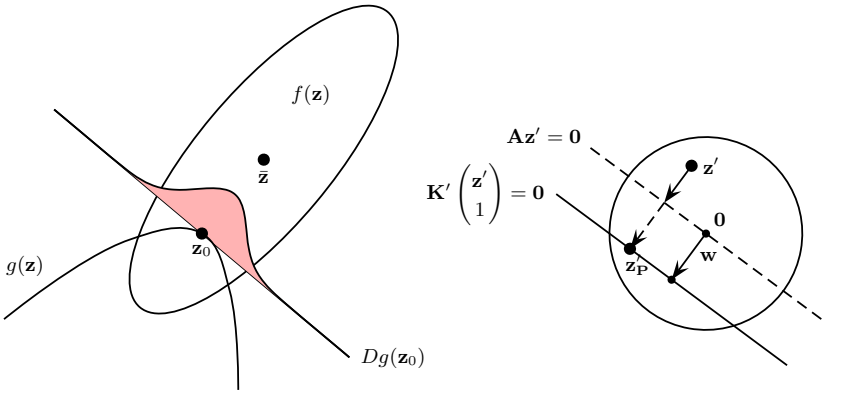
where $\bar{\mathbf{z}}$ stands for the mean pose and shape vector. In a conditional PDM, however, we additionally require that the samples coincide in the given landmarks $\mathbf{x}_K$. In other words, the samples have to meet the nonlinear equality constraint

$$g(\mathbf{z}) \equiv \mathbf{P}_K T(\mathbf{z}) - \mathbf{x}_K = \mathbf{0} \,, \tag{5}$$

where $\mathbf{P}_K$ is a projection matrix that drops all landmarks at position $\backslash K$. As depicted in Figure 2(a), the samples from the conditional PDM lie on a nonlinear manifold and are distributed according to $f(\mathbf{z})$. Our goal is to linearly approximate this manifold by the tangent space at the point $\mathbf{z}_0$ that has the highest probability under the normal distribution of the unconstrained PDM and draw samples from the normal distribution conditioned onto the tangent space. We elaborate both steps in the following subsections.

2.1 Finding $\mathbf{z_0}$

The most probable point on g with respect to f is the one closest to the mean pose and shape $\bar{\mathbf{z}}$. Thus, we find this point $\mathbf{z}_0$ by solving the following constrained quadratic minimization problem:



(a) Shape Space $\mathcal{Z}$ (b) Standardized Shape Space $\mathcal{Z}'$

Fig. 2. (a) Samples from the conditional PDM lie on the manifold that is given by the nonlinear constraint g and are distributed according to the normal distribution $p(\mathbf{z})$ with the Mahalanobis distance f. We linearly approximate this manifold by the tangent at $\mathbf{z}_0$, the point on g that is most probable with respect to f. Then we draw samples from the conditional normal distribution (in red) on this affine subspace. (b) This image illustrates how sample $\mathbf{z}'$ in the standardized space $\mathcal{Z}'$ is transformed to the affine subspace $\mathbf{K}'$. The transformed sample $\mathbf{z}'_\mathbf{P}$ is the sum of the projection of $\mathbf{z}'$ on the linear subspace $\mathbf{A}$ and the offset $\mathbf{w}$.

$$\begin{aligned} &\text{minimize} && f(\mathbf{z}) \\ &\text{subject to} && g(\mathbf{z}) = \mathbf{0} \end{aligned}$$

We propose to solve this optimization problem with a sequential quadratic programming (SQP) algorithm [1] that iteratively updates the estimate for $\mathbf{z}_0$. At each step t, we obtain a new estimate $\mathbf{z}_0^{(t+1)}$ by projecting $p(\mathbf{z})$ onto the affine subspace that corresponds to the linearized constraints at $\mathbf{z}_0^{(t)}$ and find the optimum on that space. The technical details are explained in the following.

We linearize function $T(\mathbf{z})$ at $\mathbf{z}_0$ by a Taylor expansion and set $\mathbf{z}_0 = \bar{\mathbf{z}}$ in the first iteration:

$$\begin{aligned} \mathbf{x} &\approx T(\mathbf{z}_0) + \left.\frac{\partial T}{\partial \mathbf{z}}\right|_{\mathbf{z}=\mathbf{z}_0} (\mathbf{z} - \mathbf{z}_0) \\ &= \bar{\mathbf{x}} + \mathbf{J}(\mathbf{z} - \bar{\mathbf{z}}) \,, \end{aligned} \tag{6}$$

where $\mathbf{J}$ denotes the Jacobian Matrix of T with respect to $\mathbf{z}$ and is given by

$$\mathbf{J} = \left(\tfrac{\partial T}{\partial \tilde{\mathbf{x}}} \; \tfrac{\partial T}{\partial \tilde{s}} \; \tfrac{\partial T}{\partial \tilde{\alpha}} \; \tfrac{\partial T}{\partial \tilde{t}_x} \; \tfrac{\partial T}{\partial \tilde{t}_y}\right)$$

Then the first order Taylor expansion for constraint function g follows from (5) and (6)

$$\begin{aligned}\mathbf{0} &= \mathbf{P}_K T(\mathbf{z}) - \mathbf{x}_K \\ &\approx (\mathbf{x}_0)_K + \mathbf{J}_K(\mathbf{z} - \mathbf{z}_0) - \mathbf{x}_K \end{aligned} \tag{7}$$

or, equivalently,

$$\mathbf{K}\begin{pmatrix}\mathbf{z}\\ 1\end{pmatrix} = 0\,,$$

where matrix $\mathbf{K}$ models the known shape part $\mathbf{x}_K$ as linear constraints and is defined by

$$\mathbf{K} \equiv \Big(\mathbf{J}_K - (\mathbf{x}_K - (\mathbf{x}_0)_K + \mathbf{J}_K\mathbf{z}_0)\Big)\ . \tag{8}$$

To compute the most probable point on the affine subspace $\mathbf{K}$, we look at the transformed coordinates where $\mathbf{z}$ is whitened.

$$\begin{aligned} f(\mathbf{z}) &= (\mathbf{z} - \bar{\mathbf{z}})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{z}}^{-1}(\mathbf{z} - \bar{\mathbf{z}}) \\ &= (\mathbf{z} - \bar{\mathbf{z}})^{\mathrm{T}}\mathbf{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Lambda}^{-1/2}\mathbf{V}^{\mathrm{T}}(\mathbf{z} - \bar{\mathbf{z}}) \\ &= \mathbf{z}'^{\mathrm{T}}\mathbf{z}'\ , \end{aligned} \tag{9}$$

where $\mathbf{z}' \in \mathcal{Z}'$ is given by

$$\mathbf{z}' \equiv \boldsymbol{\Lambda}^{-1/2}\mathbf{V}^{\mathrm{T}}(\mathbf{z} - \bar{\mathbf{z}})\ . \tag{10}$$

We call $\mathcal{Z}'$ the standardized shape space.

Using homogeneous coordinates, we can refomulate (9) as

$$\mathbf{S}\begin{pmatrix}\mathbf{z}\\ 1\end{pmatrix} = \begin{pmatrix}\mathbf{z}'\\ 1\end{pmatrix} \tag{11}$$

with

$$\mathbf{S} \equiv \begin{pmatrix}\boldsymbol{\Lambda}^{-1/2}\mathbf{V}^{\mathrm{T}} & -\boldsymbol{\Lambda}^{-1/2}\mathbf{V}^{\mathrm{T}}\bar{\mathbf{z}}\\ \mathbf{0}^{\mathrm{T}} & 1\end{pmatrix}\ . \tag{12}$$

Thus, matrix $\mathbf{S}$ transforms a sample $\mathbf{z}$ from $\mathcal{N}(\bar{\mathbf{z}}, \boldsymbol{\Sigma}_{\mathbf{z}})$ to the standardized normal distribution $\mathcal{N}(\mathbf{0}_D, \mathbf{I}_{D\times D})$, where D denotes the number of chosen pose and shape eigenvectors from the PDM. Conversely, the following inverse transformation maps $\mathbf{z}'$ from $\mathcal{Z}'$ to $\mathcal{Z}$

$$\mathbf{S}^{-1} = \begin{pmatrix}\mathbf{V}\boldsymbol{\Lambda}^{1/2} & \bar{\mathbf{z}}\\ \mathbf{0}^{\mathrm{T}} & 1\end{pmatrix}\ . \tag{13}$$

Thus, the linear transform $\mathbf{K}$ takes the following form in $\mathcal{Z}'$:

$$\begin{aligned}\mathbf{0} &= \mathbf{K}\begin{pmatrix}\mathbf{z}\\1\end{pmatrix}\\ &= \mathbf{K}\mathbf{S}^{-1}\mathbf{S}\begin{pmatrix}\mathbf{z}\\1\end{pmatrix}\\ &\underset{(11)}{=} \mathbf{K}'\begin{pmatrix}\mathbf{z}'\\1\end{pmatrix}\ ,\end{aligned} \tag{14}$$

where we define $\mathbf{K}' \equiv \mathbf{K}\mathbf{S}^{-1}$. Suppose $\mathbf{K}'$ is split into two parts

$$\mathbf{K}' = \begin{pmatrix}\mathbf{A}\ \mathbf{b}\end{pmatrix}\ , \tag{15}$$

where matrix $\mathbf{A}$ captures scaling and rotation information, while vector $\mathbf{b}$ contains the translation information of $\mathbf{K}'$. Then we obtain the most probable point on $\mathbf{K}'$ by simply computing the offset $\mathbf{w} = A^{\dagger}\mathbf{b}$, as illustrated in Figure 2(b). If we transform this point back to shape space $\mathcal{Z}$, we conclude an iteration of the SQP algorithm and obtain a new value for $\mathbf{z}_0$ in the Taylor expansion of 6.

2.2 Drawing Samples

As soon as we have found $\mathbf{z}_0$, the most likely point satisfying the nonlinear constraints, drawing samples from the tangent space is straightforward. By drawing samples in the whitened space, we proceed as follows:

1. Draw samples from the standardized normal distribution
$$\mathbf{z}'^{(l)} \sim \mathcal{N}(\mathbf{0}_D, \mathbf{I}_{D\times D})\ . \tag{16}$$
2. Project the samples with matrix $\mathbf{P} = \mathbf{I} - \mathbf{A}^{\dagger}\mathbf{A}$ on the the linear subspace $\mathbf{A}\mathbf{z}' = \mathbf{0}$
$$\mathbf{z}'^{(l)}_{*} = \mathbf{P}\mathbf{z}'^{(l)}\ . \tag{17}$$
3. Add $\mathbf{w} = \mathbf{A}^{\dagger}\mathbf{b}$ to the samples to account for the offset of the affine subspace $\mathbf{K}'\begin{pmatrix}\mathbf{z}'\\1\end{pmatrix} = \mathbf{0}$

$$\begin{aligned}\mathbf{z}'^{(l)}_{\mathbf{P}} &= \mathbf{z}'^{(l)}_{*} + \mathbf{w}\\ &= (\mathbf{I} - \mathbf{A}^{\dagger}\mathbf{A})\mathbf{z}'^{(l)} + \mathbf{A}^{\dagger}\mathbf{b}\ .\end{aligned} \tag{18}$$

Finally, we transform $\mathbf{z}'^{(l)}_{\mathbf{P}}$ back to the original basis in $\mathcal{Z}$ by applying the inverse transform $\mathbf{S}^{-1}$:

$$\begin{pmatrix}\mathbf{z}^{(l)}_{\mathbf{P}}\\1\end{pmatrix} = \mathbf{S}^{-1}\begin{pmatrix}\mathbf{z}'^{(l)}_{\mathbf{P}}\\1\end{pmatrix}. \tag{19}$$

The complete algorithm is presented in pseudocode notation at the end of this paper.

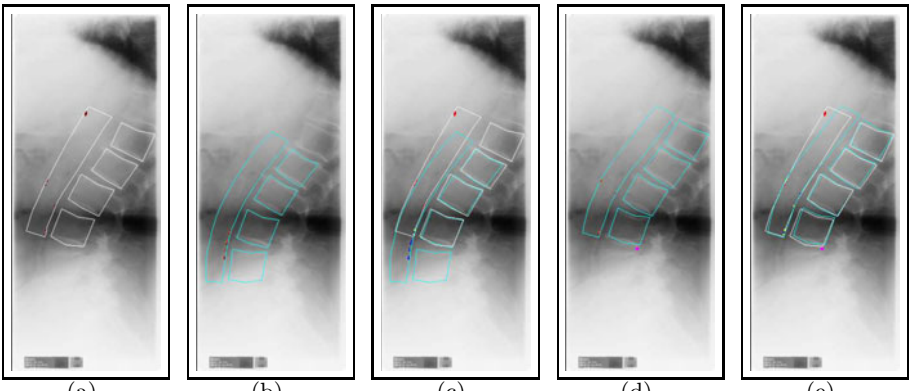


Fig. 3. This image sequence illustrates that drawing samples from a conditional PDM can be effective to resolve ambiguities in the segmentation result. In this case, the aorta and vertebrae segmentation method of [6] is constrained by a single vertebrae landmark point. (a) The manual vertebrae, aorta and calcification annotations from a medical expert. (b) The displaced conditional vertebrae mean from [6] using a PDM. (c) The overlay of the first two images. (d) The effect of constraining the search space by the landmark in the down right corner of L4. (e) The overlay of the manual annotation and the conditional PDM result. Calcifications in the overlay images are colored as follows: Yellow denotes true positives, red false negatives and blue false positives.

3 Evaluation

We evaluate our approach by comparing two configurations of the vertebrae and aorta segmentation algorithm in [6]: The first one draws samples from an unconditioned PDM, while the second one considers a manual point on the vertebrae boundary in a conditional PDM. More specifically, the conditional PDM is evaluated on lateral spine radiographs taken from a combined osteoporosis-atherosclerosis screening program. The dataset is scanned at a resolution of 45 μm per pixel and contains both healthy and fractured vertebrae as well as aortas with no till many visible calcifications. A medical expert outlined the aorta and placed six landmark points on each of the lumbar vertebrae L1 to L4.

In the first setup, the automated method of [6] is applied to 135 images in a 5-fold cross validation for segmenting the aorta and the vertebrae L1 to L4. The task is challenging for at least two reasons: First, the aorta is invisible and can only be indirectly inferred from position and shape of the vertebrae and potential calcifications. And second, it is difficult to assess the correct vertical position of the L1 to L4 vertebrae. The chosen segmentation method displaces the vertebrae in 35 out of the 135 images (about 26%). This affects the overall performance of the segmentation algorithm, but is commensurate with the number of shifts reported in [7].

In the second setup, we placed one point on the vertebae boundary in each of the 35 images with a displaced vertebrae and quantified how much the conditional

mean of the final aorta and vertebrae samples deviates from the manual annotation averaged over the landmark points:

	Vertebrae Distance [mm]	Aorta Distance [mm]	Aorta Overlap
PDM	6.87 (± 1.12)	9.43 (± 12.83)	0.48 (± 0.07)
Conditional PDM	1.38 (± 0.54)	3.18 (± 2.66)	0.73 (± 0.10)

Figure 3 depicts the effect of constraining a PDM by a single vertebrae landmark. We observe that the obtained vertebrae sample does not exactly hit the fixed landmark. The reason is that the proposed conditional PDM is based on a linear approximation to the nonlinear manifold g.

However, a second landmark forces the samples much closer to g. We evaluated how the number and position of landmarks influences the mean distance from the fixed point to the respective landmark points of the samples (see also Figure 4). The next table states the error in [mm] for different number of landmarks based on 1000 samples averaged over the 35 images. Columns 2-6 show the distance using a PDM conditioned on no to six neighboring landmarks on vertebra L4, whereas columns 7-8 report the distance for more distributed landmarks.

	0	1	2	3	6	2 (apart)	3 (apart)
mean	44.40	1.45	0.36	0.16	0.03	0.09	0.03
std	26.98	1.73	0.38	0.15	0.03	0.09	0.03

4 Discussion

We demonstrated how we can effectively constrain the solution space of sample-based segmentation methods that rely on a PDM. As in the vertebrae segmentation example, we can refine the solution by creating a PDM that is conditioned on one or two given landmark points. As we have seen, the overall performance of the segmentation algorithm might improve considerably.

In this paper, we modeled linear constraints to approximate the general procrustes alignment in the PDM. The samples might therefore not exactly meet the fixed point, but be very close to it. We are currently investigating how to sample from the nonlinear manifold g exactly.

Another promising extension of this technique concerns the automated creation of constraint points. So far, a human expert was responsible for placing landmark points. However, we might also constrain the PDM in sampling-based segmentation by automatically determining salient points that constrain the position and shape of an object. For instance, the recognition of likely calcifications could help detect plausible aortas and be an important step towards quantifying calcifications for the prognosis and diagnosis of CVD and mortality.

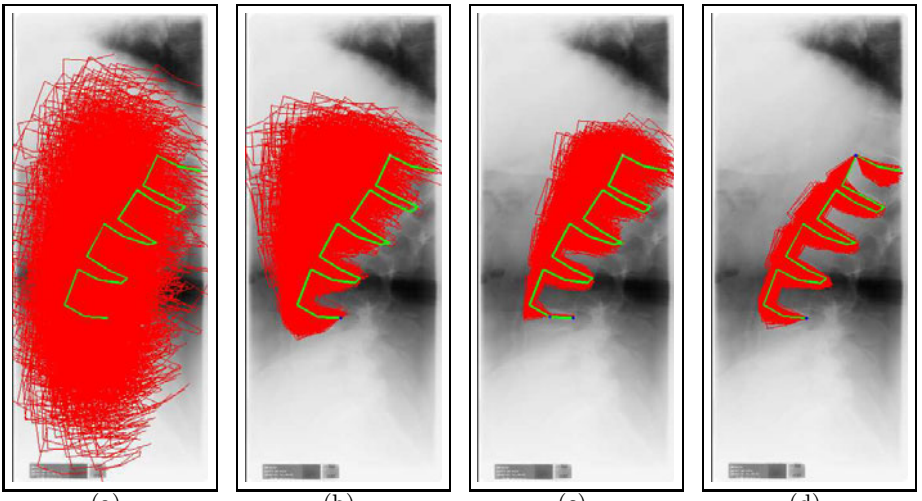


Fig. 4. (a) Vertebrae samples (red) of method [6] without constraints and the manual vertebrae annotation in green. (b) The sample space if the bottom right point is constrained (blue point) in a conditional PDM. (c) Two close-by constraints. (d) Two distanced constraint points.

References

1. Bertsekas, D.P.: Nonlinear Programming, 2nd edn. Athena Scientific (1999)
2. De Bruijne, M.: Shape particle guided tissue classification. In: Golland, P., Rueckert, D. (eds.) Mathematical Methods in Biomedical Image Analysis, MMBIA (2006)
3. de Bruijne, M., Nielsen, M.: Image segmentation by shape particle filtering. In: ICPR, Washington, DC, USA, pp. 722–725. IEEE Computer Society, Los Alamitos (2004)
4. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models – their training and application. Computer Vision and Image Understanding 61(1), 38–59 (1995)
5. Gower, J.: Generalized procrustes analysis. Psychometrika 40(1), 33–51 (1975)
6. Petersen, K., Nielsen, M., Brandt, S.S.: A Static SMC Sampler on Shapes for the Automated Segmentation of Aortic Calcifications. In: Daniilidis, K. (ed.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 666–679. Springer, Heidelberg (2010)
7. Roberts, M.G., Cootes, T.F., Pacheco, E., Oh, T., Adams, J.E.: Segmentation of Lumbar Vertebrae Using Part-Based Graphs and Active Appearance Models. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 1017–1024. Springer, Heidelberg (2009)

Algorithm 1. Sampling from Conditional Point Distribution Model (PDM)

Require: a PDM with mean shape and pose parameters $\bar{\mathbf{z}} = (\tilde{\bar{\mathbf{x}}}, \bar{\theta})$, covariance matrix $\mathbf{\Sigma_z}$ with $\mathbf{z} = (\theta, \tilde{\mathbf{x}})$, and shape to measurement space transform $T(\mathbf{z})$; a vector of known points $\mathbf{x}_K$ (constraint indices K); a convergence threshold $\epsilon \in \mathbb{R}^+$.

Ensure: L samples $\mathbf{z}_\mathbf{P}^{(l)}$ from the conditional PDM.

Initialization:
Compute inverse transform $\mathbf{S}^{-1}$ for moving to standardized shape space (13)
$t \leftarrow 0$

Optimization of Expansion Point $\mathbf{z}_0$ (SQP Algorithm):
repeat
  if $t = 0$ **then**
    $\mathbf{z}_0^{(t)} \leftarrow \bar{\mathbf{z}}$
  else
    $\mathbf{z}_0^{(t)} \leftarrow \mathbf{z}_0^{(t+1)}$
  end if
  Compute Jacobian matrix $\mathbf{J}$ at expansion point $\mathbf{z}_0^{(t)}$ for constraint function $g(\mathbf{z})$.
  Compute constraint matrix

$$\mathbf{K} \leftarrow \left(\mathbf{J}_K - \left(\mathbf{x}_K - \bar{\mathbf{x}}_K + \mathbf{J}_K \mathbf{z}_0^{(t)}\right)\right) , \qquad \bar{\mathbf{x}}_K = \mathbf{P}_K T(\bar{\mathbf{z}})$$

  .
  Determine $\mathbf{A}$ and $\mathbf{b}$ from $\mathbf{K}' \leftarrow \mathbf{K}\mathbf{S}^{-1}$ (15).
  Update expansion point

$$({\mathbf{z}_0}^{(t+1)}, 1)^{\mathrm{T}} \leftarrow \mathbf{S}^{-1}\mathbf{A}^\dagger \mathbf{b} .$$

until $|\mathbf{z}_0^{(t+1)} - \mathbf{z}_0^{(t)}| < \epsilon$.

Sampling:
for $l = 1$ to L **do**
  Sample $\mathbf{z}'^{(l)}$ from standardized normal distribution (16).
  Project $\mathbf{z}'^{(l)}$ to affine subspace

$$\mathbf{z}_\mathbf{P}'^{(l)} \leftarrow (\mathbf{I} - \mathbf{A}^\dagger \mathbf{A})\mathbf{z}'^{(l)} + \mathbf{z}_0^{(t+1)} .$$

  Obtain sample $\mathbf{z}_\mathbf{P}^{(l)}$ by transforming $\mathbf{z}_\mathbf{P}'^{(l)}$ back to the original shape space (19).
end for

Deformable Registration of Organic Shapes via Surface Intrinsic Integrals: Application to Outer Ear Surfaces

Sajjad Baloch, Alexander Zouhar, and Tong Fang

Siemens Corporate Research, Princeton, NJ, USA

Abstract. We propose a method for the deformable registration of organic surfaces. Meaningful correspondences between a source surface and a target surface are established by means of a rich surface descriptor that incorporates three categories of features: (1) local and regional geometry; (2) surface anatomy; and (3) global shape information. First, surface intrinsic, *geodesic distance integrals*, are exploited to constrain the global geodesic layout. Consequently, the resulting transformation ensures topological consistency. Local geometric features are then introduced to enforce local conformity of various regions. To this end, the extrema of appropriate curvatures – the extrema of mean curvature, minima of Gauss and minimum principal curvature, and the maxima of maximum principal curvature – are considered. Regional features are introduced through *curvature integrals* over various scales. On top of this, *explicit anatomical priors* are included, thereby resulting in anatomically more consistent registration. The source surface is deformed to the target by minimizing the energy of matching the source features to the target features under a Gaussian propagation model. We validate the proposed method with application to the outer ear surfaces.

1 Introduction

The registration of organic surfaces is a challenging problem and far more complex to be adequately addressed solely by the geometric information. Anatomical correspondence may not be uniquely determined from the geometric attributes; or it may not exist at all due to anatomical variability. For instance, it may not be possible to perfectly warp a single-fold sulcus to a double-fold through a biologically meaningful transformation. Or, the concha may not be prominent in the registration of outer ear surfaces. Furthermore, extra material may also deceive the algorithm by creating false anatomy like structures. In such cases, geometry alone is insufficient, and our hypothesis is that richer representations are required that combine anatomy, shape as well as regional and local geometry.

Various methods have appeared in literature. Basic geometric entities such as surface points [3], local shape information in so-called spin-maps [7], spherical harmonics [9] as well as 3D shape contexts [15] have been proposed for rigid registration. Recently, [14] proposed a combination of anatomical and geometric features for anatomically aware registration of ear surfaces.

B. Menze et al. (Eds.): MICCAI 2010 Workshop MCV, LNCS 6533, pp. 11–20, 2011.

For non-linear registration, coarse-to-fine scale curvature based features [5], average convexity features [6], and mean curvatures [13] have been employed. Most of these features are local in nature and do not perform well when anatomy is not adequately represented by geometric primitives. To overcome this limitation, shape information was incorporated via intrinsic shape contexts (ISC) [11] or the clamp histograms [10] with promising results.

A major limitation of these methods is that they do not generalize well for arbitrary surfaces. For example, the dependence of clamp histograms solely on surface normals makes them less attractive for surfaces with somewhat flat patches. Similarly, ISC does not exhibit much variability between equally spaced histogram bins due to geodesic binning, and fails to differentiate among concave, convex, and saddle regions. Furthermore, it does not perform well for surfaces with boundaries or in the presence of excess material. On the other hand, 3D shape contexts do not know anything about the topology of a surface, and are susceptible to topologically incorrect correspondences.

We propose a very rich descriptor for surface representation that is intrinsic[1] to the surface. It captures the global shape and the geodesic layout of a surface through *geodesic distance integrals* (GDI). This ensures that the registration is spatially more consistent thereby allowing large deformations without distorting topological structure of a surface. The global shape is then coupled with regional and local *curvature integrals* evaluated at various scales. To this end, we consider (a) the maxima and minima of the mean curvature, (b) the minima of the minimum principal curvature, (c) the maxima of the maximum principal curvature, and (d) the minima of the Gauss curvature. Consequently, concavities are matched to concavities, convexities are matched to convexities, and saddles are matched to saddles on the two surfaces. Introduction of various scales not only helps in avoiding local minima leading to better optimization of the underlying energy functional, but also takes into account the relative sizes of the convex, concave and saddle regions. On top of this shape and geometric information, explicit anatomical priors are introduced for anatomically more plausible registration. The smoothness constraints are enforced by a Gaussian propagation model [10]. Most of the aforementioned existing work was tailored to brain data. Our richer formulation makes our method potentially more suited for wide range of organic surfaces. It does not require surface to be closed, or highly convoluted.

We validate our method with application to the registration of outer ear surfaces [12]. The application is significant for the construction of a 3D digital atlas of the human ear, as well as for hearing aid manufacturing. For the anatomical prior, we capitalize on the canonical ear signature (CES) [2].

2 Problem Formulation

Given a surface $\mathcal{M}_S$ representing the source anatomy and a surface $\mathcal{M}_T$ representing the target anatomy, the problem under consideration is to estimate a

[1] Intrinsic in the sense that it does not depend on the orientation of the surface.

diffeomorphic transformation $h : \mathcal{M}_S \rightarrow \mathcal{M}_T, \mathcal{M}_S \mapsto h(\mathcal{M}_S)$ that warps $\mathcal{M}_S$ to $\mathcal{M}_T$, by minimizing the bending energy:

$$E(h) := \omega_e E_e(h) + \omega_i E_i(h),$$

where E_e is the external energy term that depends on the source and target surfaces, and E_i is the internal energy term that ensures the smoothness of the transformation. For the results in this paper, h is modeled by a Gaussian propagation similar to [10], although other approaches may be utilized [4]. ω_e and ω_i determine the relative importance of the two terms.

3 External Energy

We are interested in the external energy that establishes correspondence between the source and target surfaces by identifying key surface landmarks. To this end, we consider geometric and anatomical landmarks. E_e takes the form:

$$E_e(h) := \gamma_G E_e^G(h) + \gamma_F E_e^F(h), \tag{1}$$

where γ_G, γ_F are the weights, and E_e^G and E_e^F denote the geometric and anatomical components respectively. The former is derived from a rich surface descriptor described next and in turn has two components (the global shape $E_e^S(h)$ and the local/regional geometry $E_e^R(h)$), whereas the latter involves anatomical features.

3.1 Proposed Surface Descriptor

The geometric component of the bending energy ensures the establishment of correspondences between geometrically similar regions on the two surfaces, based on certain key geometric landmarks. To this end, we propose a rich surface descriptor that takes into account the local, regional, and global geometric and shape information. Since this descriptor is defined on both surfaces in a similar way, we drop the subscripts S and T in the next few sections.

Global Information. For global shape and topology, *geodesic distance integrals* (GDI) are defined on the surface. The idea is to ensure that the source-target correspondences respect their respective global layouts. A GDI at a point $u \in \mathcal{M}$ is defined as the integral of its geodesic distance $g(u, x)$ $\forall x \in \mathcal{M}$:

$$\mathcal{S}(u) := \int_{x \in \mathcal{M}} g(u, x) d\mathcal{M}.$$

A GDI is a *Morse* function [1], and its critical points are important topological landmarks. Like *spin-maps* [7], GDIs are intrinsic to a surface. On the other hand, they have a global support region in contrast to the spin-maps. Introduction of the GDI ensures that the resulting correspondences respect their global geodesic layouts, and the deformations that disturb the geodesic arrangement of points are discouraged. This in particular is very powerful for large scale

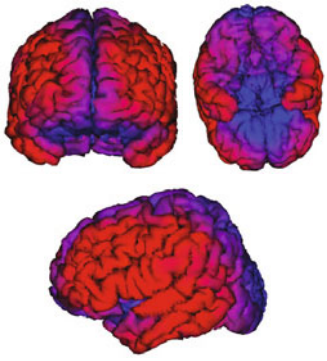


Fig. 1. Normalized GDI maps for a representative human brain (values range from blue for 0 to red for 1)

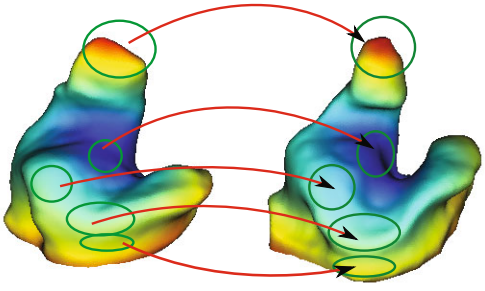
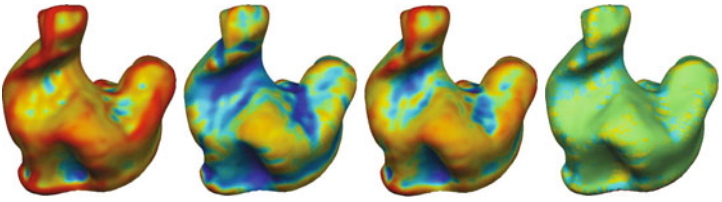


Fig. 2. Normalized GDI maps (values range from blue for 0 to red for 1):(Left) Source surface; (Right) Target surface. Mapped circles indicate layout-wise similar regions. Small values occur in the middle of a surface, whereas large values occur at regions which are most distant from the rest of the surface (e.g., the canal tip).

deformations, and allows one to greatly increase the search space without getting stuck in a local minimum. The GDIs for a human brain are illustrated in Fig. 1, and that for two arbitrary ear surfaces is given in Fig. 2, along with the correspondences.

Local Information. While GDIs maintain the higher level shape deformation, most local variations are controlled through localized features. Here, we use the extrema of the mean curvature, κ_μ, as a feature of interest. However, mean curvature alone is not sufficient for all possible local deformations. We, therefore, include the minima of the minimum principal curvature, κ_{pc_1}, and the maxima of maximum principal curvature, κ_{pc_2}, to ensure radial concavity-concavity and radial convexity-convexity correspondences. The minima of Gauss curvature, κ_G, is also considered due to its ability to capture saddle points. The feature vector at a point $u \in \mathcal{M}$ is, therefore, composed of $\boldsymbol{\alpha}_l(u) : (\kappa_\mu(u), \kappa_G(u), \kappa_{pc_1}(u), \kappa_{pc_2}(u))$. At this point, the individual components are not the extrema of curvatures, which will be done later via landmark selection.

Regional Information. Note that the curvature is a local measure of bending, and is not appropriate while matching geometric landmarks of various sizes. For instance, if a bump on a surface has to be matched to two candidate bumps on a target surface, one has to resolve the ambiguity. In such a case, the size of the bump plays an important role. For this reason, we compute *curvature integrals* at various scales. An integral of the above feature vector in a geodesic neighborhood of size s yields a scale space representation of the vector $\boldsymbol{\alpha}_r(u; s)$. Our geometric feature vector is, therefore, defined as: $A(u) := [\boldsymbol{\alpha}_r(u; s_1), \ldots, \boldsymbol{\alpha}_r(u; s_k), \mathcal{S}(u)]$.



(a) (b) (c) (d)

Fig. 3. Normalized curvature maps – values range from low (blue) to high curvatures (red): (a) Mean curvature; (b) Minimum principal curvature; (c) Maximum principal curvature; (d) Gauss Curvature. Note that extrema of mean curvature, minima of minimum principal curvature and Gauss curvature, and maxima of maximum principal curvature correspond to geometrically significant landmarks.

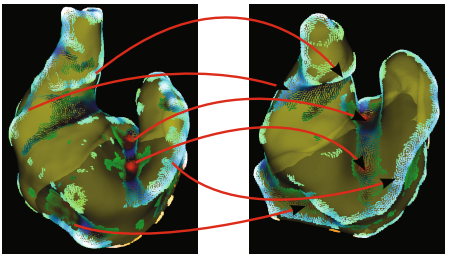


Fig. 4. Geometric landmarks on ear surfaces: (Left) Source; (b) Target. Arrows indicate the correspondences. The landmarks are color coded according to their curvature profile, where $(\kappa_G, \kappa_{pc_1}, \kappa_{pc_2})$ are set as the *rgb*-components.

Note that A does not explicitly include $\boldsymbol{\alpha}_l$; it is implicitly included since $\boldsymbol{\alpha}_l$ is a special case of $\boldsymbol{\alpha}_r$, and by varying s from small to a large value, one can capture geometric information from local to regional level. Note that all features are normalized to the range [0,1] before combining them.

3.2 Surface Landmarks

The first term in Eq. (1) is composed of global shape E_e^S and local/regional geometric E_e^R components. To ensure global layout coherence all surface points are considered in E_e^S. E_e^R, on the other hand, establishes correspondences between geometrically interesting points. To this end, we rely on local extrema of mean curvature, minima of Gauss and minimum principal curvature, maxima of maximum principal curvature. All points exhibiting local extrema are more significant and are therefore, regarded as *landmarks*. In short, our registration algorithm is driven by two kinds of landmarks: (1) anatomical landmarks, and (2) geometric landmarks along with the global shape.

Instead of selecting just the local extrema, we control the geometric landmark selection via thresholding. A set of points $\mathcal{D} \subset \mathcal{M}$ is constructed from points exhibiting more "curvedness" than the prescribed thresholds. Later, thresholds are gradually relaxed in the course of algorithm evolution to gradually put more emphasis on less interesting points. Geometric landmarks for an arbitrary shape at some arbitrary threshold are shown in Fig. 4. Note that GDIs require a different treatment as mentioned above.

Fig. 5. Optimization strategy: $\mathcal{M}_S$ is first rigidly registered to $\mathcal{M}_T$, followed by anatomical, global layout, and local/regional geometric deformations

4 Minimization of Bending Energy

Minimization of E is carried out in blocks according to Fig. 5, where each component is of the form: $\sum ||h(\ell_S^i) - \ell_T^i||_2$ (ℓ is landmark of interest). First, rigid alignment is carried out, followed by the deformation under anatomical constraints. Finally, global layout and regional features are considered. Each deformation block consists of the following steps, where h is initialized with the output of the previous block, and updated iteratively until the energy drop becomes negligible.

1. At each iteration k, landmarks on the source surface are identified. For GDIs, it means the entire surface, and for curvature integrals, this amounts to values greater than a certain threshold.
2. Each landmark, $u_s \in \mathcal{D}_S$, is mapped to $u_t \in \mathcal{M}_T$ in the following manner. First, $\mathcal{D}_S$ is deformed under h^k to yield the set $h^k(\mathcal{D}_S)$. Each $h^k(u_s) \in h^k(\mathcal{D}_S)$ is then mapped to $u_T \in \mathcal{M}_T$ through closest point projection. A neighborhood region $\mathcal{N}_T(u_T; r)$ of size r is selected around u_T. Finally, correspondence is established by finding the best match according to: $v* = \arg\min_{v \in \mathcal{N}_T(u_T;r)} ||A_S(u_s) - A_T(v)||$. This step is skipped for anatomical deformation, since correspondences are explicitly defined.
3. The corresponding points define the displacements $d^k = \mathcal{M}_T(v*) - \mathcal{M}_S(u)$ of landmark points.
4. In one iteration, we are interested in only a small ($\delta > 0$) step towards the target surface. The differential displacement (δd^k) is propagated over the entire surface.
5. Each surface point is deformed by the corresponding differential displacements $h^{k+1}(u) = h^k + \delta d^k(u), \forall u \in \mathcal{M}_S$.

For the results in this paper, we considered only one pass through the block diagram. One may also opt for repeated passes of the last three blocks.

5 Application to Ear Surfaces

So far, the proposed method has been developed in a general setting. We now apply it for the registration of ear surfaces (Fig. 6). Application to other organic shapes does not require any changes in the algorithm. Only the prior anatomical information will be modified. Alternatively, one may select $\gamma_F = 0$ in Eq. (1) to drive the algorithm solely by geometric information.

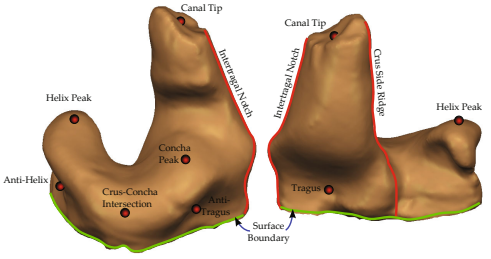


Fig. 6. Anatomical features on the human ear

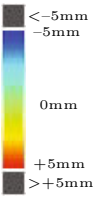


Fig. 7. Color bar for the registration error maps. A light green map means ideal registration.

For anatomical features, we utilize the recently proposed *canonical ear signature* (CES), which is a comprehensive representation of the human ear anatomy. In short, CES consists of anatomical points, contours, regions, and planes. Among them, features of particular interest are *canal tip point*, *helix peak point*, *concha peak point*, *tragus point*, *anti-tragus point*, *anti-helix point*, *center crus concha*, *inter-tragal notch*, and *crus-side canal ridge*, shown in Fig. 6. These features can be robustly and efficiently detected fully automatically [2].

6 Experiments

We conducted a series of experiments for the registration of ear surfaces, as well as for label transfer. Ear surfaces were acquired from digital scans of 3D ear impressions taken from 17 subjects. One of the surfaces was randomly selected as template. All subjects were *rigidly* aligned to the template using [14].

16 remaining surfaces were automatically registered to the template using the algorithm outlined above. It involved automatic detection of anatomical features on all subjects including the template. Scales were chosen as $s \in [0.5, 3]$ in steps of 0.5; thresholds for geometric landmarks were selected as the mean $\pm$ twice the standard deviation of respective curvatures, and were gradually relaxed by a fraction of 1.1.

Comparison. A comparison of the proposed method with [14] is provided in Fig. 8. The results indicate excellent registration, where the helix, the canal, and the crus area are matched very nicely. For [14], they were more than the "outlier" distance away from the target shape (shown by the metal color in Fig. 8(c)). After registration, the error map is uniformly light green (Fig. 8(d)).

Quantitative Validation. An expert was asked to manually segment the template, who provided coarse as well as detailed labelings (Fig. 9). These labelings (GT) were then used for the segmentation of the test shapes (Fig. 10). The results are in accordance with the underlying geometry. Notice the boundary between green and yellow passes right through the valley. The canal is also perfectly

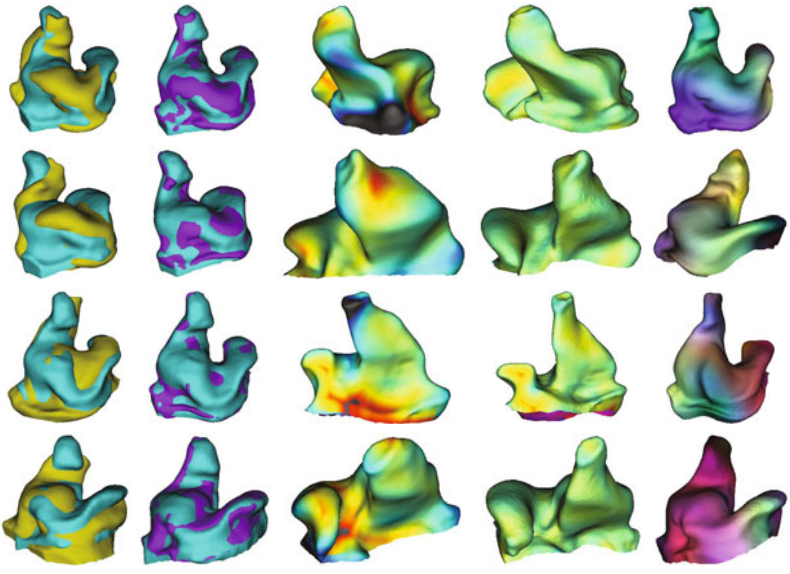


Fig. 8. Deformable registration of representative surfaces: (a) Subject registered with [14] (gold), template (cyan); (b) Registered by proposed method (purple); (c) Error map before registration (Color bar in Fig. 7); (e) Error map after registration; (f) Diffeomorphism h on rgb-scale

segmented. Labeling was quantitatively evaluated for each label as ($\text{Area}_{\text{label}} \cap \text{Area}_{\text{GT}}$) / ($\text{Area}_{\text{label}} \cup \text{Area}_{\text{GT}}$), with an average measure of 0.94.

Evaluation of Surface Descriptor. We evaluated the effect of various components of our surface descriptor. Three cases were considered: (1) registration based solely on anatomical information; (2) combined anatomy and GDIs; and (3) the complete surface descriptor along with the anatomical information. Results are presented in Fig. 11. Notice that the complete descriptor (anatomy + GDIs + Local/Regional features) yields the best performance. In Fig. 12, the labeling as a result of cases (2) and (3) are presented. Notice how the complete descriptor in Fig. 12(Right) removes problems highlighted in Fig. 12(Left), such

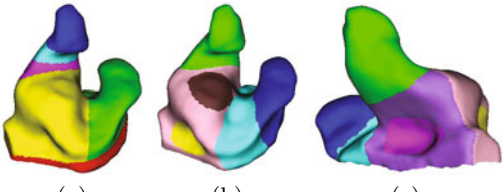


Fig. 9. Expert manual segmentation of the template: (a) Coarse; (b)-(c) Detailed

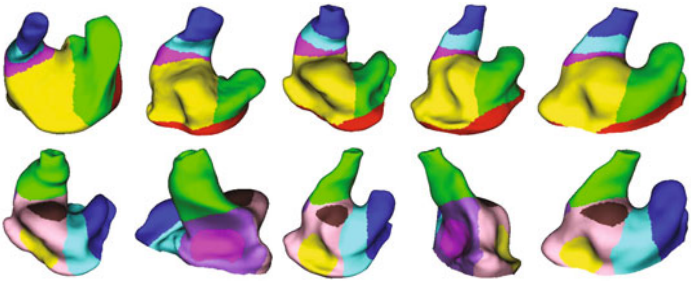


Fig. 10. Label transfer from the template based on estimated diffeomorphism: (Top Row) Coarse Segmentation; (Bottom Row) Detailed Segmentation

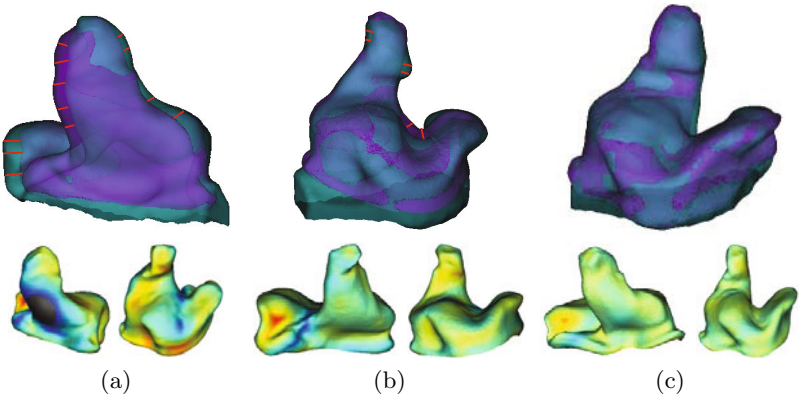


Fig. 11. Evaluation of various features: (a) Anatomical features; (b) Anatomy + GDI; (c) Complete descriptor. (Top) Warped surfaces; Red lines indicate regions, where the registration does not perform well. (Bottom) Error map (Color bar in Fig. 7). Complete descriptor yields almost uniform error map.

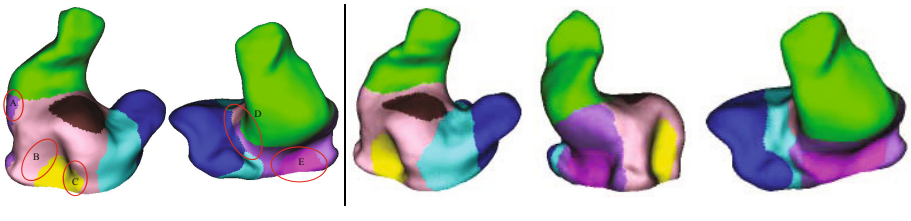


Fig. 12. Labeling based on (Left) Anatomy + GDI; (Right) Complete descriptor, which fixes the highlighted regions

as over-segmentation in regions A and C, and under-segmentations in regions B and E are gone. In particular, in region D, the anatomy (cyan color) was split into two parts, which is now resolved.

7 Conclusions

We have proposed a method for automatic registration of organic surfaces. The method incorporates anatomical priors for anatomically consistent correspondences. It also introduces a novel feature vector that is more distinguishing than the existing ones. To this end, geodesic distance integrals were used for global and curvature integrals at various scales are used for regional and local information. The method has been applied for the registration and label transfer of ear surfaces. In future, we plan to construct a comprehensive digital atlas of the ear anatomy using the proposed method and the anatomical signature [2].

References

1. Aouada, D., Dreisigmeyer, D., Krim, H.: Geometric modeling of rigid and non-rigid 3D shapes using the global geodesic function. In: CVPR Workshops, pp. 1–8 (2008)
2. Baloch, S., et al.: Automatic detection of anatomical features on 3D ear impressions for canonical representation. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6363, pp. 555–562. Springer, Heidelberg (2010)
3. Besl, P.J., McKay, N.D.: A method for registration of 3-D shapes. IEEE PAMI 14(2), 239–256 (1992)
4. Bookstein, F.L.: Principal warps: thin-plate splines and the decomposition of deformations. IEEE Trans. on PAMI 11(6), 567–585 (1989)
5. Davatzikos, E.: Spatial transformation and registration of brain images using elastically deformable models. Comput Vision Image Understanding 66 (1997)
6. Fischl, B., Sereno, M., et al.: High-resolution intersubject averaging and a coordinate system for the cortical surface. Human Brain Mapping 8(4), 272–284 (1999)
7. Johnson, A.E., Hebert, M.: Surface matching for object recognition in complex three-dimensional scenes. Image & Vision Computing (16) (1998)
8. Liu, T., Shen, D., Davatzikos, C.: Deformable registration of cortical structures via hybrid volumetric and surface warping. Neuroimage 22(4), 1790–1801 (2004)
9. Makadia, A., Patterson, A., Daniilidis, K.: Fully automatic registration of 3D point clouds. In: IEEE CVPR (2006)
10. Shen, D., Davatzikos, C.: HAMMER: hierarchical attribute matching mechanism for elastic registration. IEEE TIP 21(11), 1421–1439 (2003)
11. Shi, Y., Thompson, P.M., et al.: Direct mapping of hippocampal surfaces with intrinsic shape context. Neuroimage (2007)
12. Slabaugh, G., et al.: 3D shape modeling for hearing aid design. IEEE Signal Processing Magazine 5(25) (2008)
13. Yeo, B.T.T., Sabuncu, M.R., et al.: Spherical demons: Fast diffeomorphic landmark-free surface registration. IEEE TMI 29(3), 650–668 (2010)
14. Zouhar, A., Fang, T., et al.: Anatomically-aware, automatic, and fast registration of 3D ear impression models. In: 3DPVT, pp. 240–247 (2006)
15. Zouhar, A., Baloch, S., et al.: Freeform shape clustering for customized design automation. In: IEEE 3DIM (2009)

Iterative Training of Discriminative Models for the Generalized Hough Transform

Heike Ruppertshofen[1,2], Cristian Lorenz[3], Sarah Schmidt[4,2], Peter Beyerlein[4], Zein Salah[2], Georg Rose[2], and Hauke Schramm[1]

[1] Institute of Applied Computer Science,
University of Applied Sciences Kiel, Germany
[2] Institute of Electronics, Signal Processing and Communication Technology,
Otto-von-Guericke University Magdeburg, Germany
[3] Department Digital Imaging, Philips Research Hamburg, Germany
[4] Department of Engineering, University of Applied Sciences Wildau, Germany
heike.ruppertshofen@fh-kiel.de

Abstract. We present a discriminative approach to the Generalized Hough Transform (GHT) employing a novel fully-automated training procedure for the estimation of discriminative shape models. The technique aims at learning the shape and variability of the target object as well as further confusable structures (anti-shapes), visible in the training images. The integration of the learned target shape and anti-shapes into a single GHT model is implemented straightforwardly by positive and negative weights. These weights are learned by a discriminative training and utilized in the GHT voting procedure. In order to capture the shape and anti-shape information from a set of training images, the model is built from edge structures surrounding the correct and the most confusable locations. In an iterative procedure, the training set is gradually enhanced by images from the development set on which the localization failed. The proposed technique is shown to substantially improve the object localization capabilities on long-leg radiographs.

Keywords: Object Localization, Generalized Hough Transform, Discriminative Training, Machine Learning, Optimal Model Generation.

1 Introduction

Object localization is an important prerequisite for automatic execution of many applications in computer-aided diagnosis, like, e.g., segmentation procedures. Nevertheless, only few methods for object localization exist and in many cases the localization task is still carried out manually or very task-specific solutions are developed for the localization problem at hand; frequently employing anatomical knowledge along with basic image processing methods like gray-value thresholding or morphological operators [5]. Most of these approaches require prior knowledge of the target object and its appearance, and the adaptation to new problems is usually difficult and time-consuming.

B. Menze et al. (Eds.): MICCAI 2010 Workshop MCV, LNCS 6533, pp. 21–30, 2011.

More general approaches are given by automatic localization procedures, which have the advantage that the results are reproducible and independent of the operator. Many of these methods are based on a global search of the target object employing, e.g., evolutionary algorithms [6], template matching [8], or the Generalized Hough Transform (GHT) [1,9,14]. These techniques are often time and memory consuming due to the vast search. Further interesting methods include atlas-based registration [10,15] and marginal space learning [16], which determines the object position and further parameters like scaling and rotation iteratively.

We will focus on the GHT for object localization and the models used therein. These models can either represent the shape of the object characterized by a point cloud [14,12] or the appearance described by a codebook of image patches [9,11]. To further improve the localization accuracy of their codebooks, several groups [3,11] have explored discriminative training procedures to obtain weighted models. However, none of these training approaches incorporate confusable objects to decrease false positive rates.

Recently, we have introduced an approach utilizing the GHT in combination with a discriminative training algorithm to obtain an optimized shape model directly from the data [13]. In the training procedure, the model points are assigned individual weights, which are used in the voting procedure of the GHT and to identify unimportant model points. This approach results in more discriminative models, since important points, representing the object well, will be assigned higher weights such that their impact on the object localization increases. Variations in scaling and rotation of the target object are incorporated into the model, so that in combination with slim models, the GHT becomes computationally feasible.

In this paper, we will extend the training procedure by an iterative approach, which gradually establishes a set of suitable training images, representing the variability of the target object contained in a given dataset. Manually determining such a dataset is a challenging and also crucial task, since the quality of the choice of training images is mirrored in the performance and robustness of the generated model.

Furthermore, our approach also takes rivaling structures into account, which resemble the target object. These confusable structures, also called *anti-shapes*, are identified automatically from the Hough space and are integrated into the shape model for the next training iteration. The distinction of target and anti-shape model points is realized straightforwardly through *positive* and *negative* model point weights. Through the negative weights, the model is repelled from the anti-shape structures, resulting in fewer mislocalizations and a more focused Hough space.

The method is applied here to a set of long-leg radiographs with the task to localize femur, knee and ankle. The position of these joints is needed for the initialization of a subsequent segmentation, from which angles, lengths or density of bones can be estimated. Due to varying fields of views and various artificial replacements of the hip or knee visible in the database, we cannot rely upon

prior knowledge of the object location estimated from training data [6] nor upon gray-value or appearance information [8,9,10,11,15], which is why most of the methods mentioned above are not suitable for this task. Our GHT shape based method is presented here for a 2D task, but is also applicable to 3D problems as was shown tentatively in [12].

2 Methods

2.1 Generalized Hough Transform

The GHT [1] is a method used to detect arbitrary shapes in an image. To this end the image is transformed into a parameter space, the so called Hough space, in which each cell represents a possible object location (and potentially further object transformation parameters, like rotation or scaling, which will not be considered here). The Hough space is filled via a voting procedure, where a model, representing the target object, casts votes for possible object locations. Due to the voting procedure the method is robust against image noise and occlusion or missing parts, rendering it very interesting for medical image processing.

The model used in the voting procedure consists of a point cloud representing the objects shape. The coordinates of these points are given relative to a reference point, which is usually the center (of mass) of the model. The generation of suitable models will be explained in Sec. 2.3.

To extract the shape information from the inspected image an edge operator (e.g. Canny) is applied and only the resulting edge points will be considered in the voting procedure. By assuming that the model point m_j corresponds to the edge point e_l, a possible object location is computed via $c_i = e_l - m_j$ and the vote in the related Hough cell c_i is increased. At the end of the voting procedure the cell with the highest number of votes is searched for and returned as the most likely object position.

To speed up the procedure and to suppress spurious votes, directional information is included in the model and model points are grouped accordingly in a so called R-table. Only model points which lie in the R-table bin representing the gradient direction of the current edge point will be allowed to vote.

As was pointed out previously, the GHT can be used to localize scaled and rotated occurrences of the object, as well. In this case the model is transformed and the procedure is repeated, resulting in a larger and higher dimensional Hough space. Due to performance reasons, no further parameters besides the object position will be considered here. Instead, we want to capture the underlying variability of the target object and integrate it into the model using the training and model generation procedures described in the following sections, such that no scaling or rotational information is necessary for a successful localization.

2.2 Discriminative Model Training

In the standard GHT each model point has an equal impact on the estimation of the true object position. To design the models in a more discriminative

way, an individual weight for each model point will be trained such that points robustly representing the target (rival) object will obtain a large positive (negative) weight. Points which do not assist in object localization will receive low absolute weights and can later be eliminated from the model.

To estimate the point weights, we regard each model point as individual source of knowledge, yielding information about the correct location of the object. By means of a suitable weighted combination of model point information, weights will be determined with respect to a minimal localization error on training images.

The information content of each model point is given through its contribution to the Hough space, which can be regarded as posterior probability:

$$p_j(c_i|X) = \frac{N_{i,j}}{N_j} \ , \tag{1}$$

where $N_{i,j}$ and N_j are the number of votes from model point j in the Hough cell c_i and in the complete Hough space, respectively, and X represents the features extracted from the considered image. To combine the posterior probabilities (1) from all model points a log-linear approach is employed following the Maximum-Entropy principle [7]:

$$p_\Lambda(c|X) = \frac{\exp\left(\sum_j \lambda_j \cdot \log p_j(c|X)\right)}{\sum_i \exp\left(\sum_j \lambda_j \cdot \log p_j(c_i|X)\right)} \ . \tag{2}$$

The coefficients $\Lambda = (\lambda_j)_j$ assess the influence of the posterior probabilities p_j on the model combination p_Λ and will be used as model point weights. To estimate the weights λ_j with respect to a minimal localization error [2], a smoothed error function E is defined, which accumulates the error for each of the training images X_n:

$$E(\Lambda) = \sum_n \sum_{c'} \frac{p_\Lambda(c'|X_n)^\eta}{\sum_c p_\Lambda(c|X_n)^\eta} \cdot \|c_n - c'\|_2 \ . \tag{3}$$

In the function $E(\Lambda)$, the Euclidean distance of the correct cell c_n and a cell c' in the Hough space is weighted by an indicator function, such that a large vote in a cell far away from the true solution is penalized stronger than in a cell close to the correct one. The exponent η in the indicator term regulates the influence of rivaling hypotheses c on the error measure.

For the estimation of weights λ_j from (3) a gradient descent scheme is explored. Due to the high-dimensional search space and the most likely non-convex error surface, finding a global minimum cannot be guaranteed. Nevertheless, the λ_j resulting from a local minimum already show a clear improvement in localization accuracy as can be seen in Sec. 3.

2.3 Model Generation and Design of Experiments

To be able to exploit the strength of the training technique, suitable initial models need to be created, which allow for the integration of shape variability

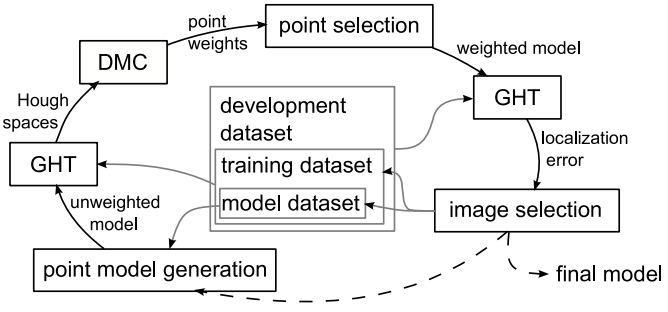


Fig. 1. Schematic overview of the iterative training procedure for creation of discriminative models as described in Sec. 2.3. The procedure makes use of the GHT and the discriminative training procedure (DMC), described in the previous sections.

and anti-shapes. Another aim is to obtain a fully automatic system that learns a model purely from the data without the incorporation of expert knowledge or user interaction.

To this end an iterative procedure is introduced, employing a set of annotated development images with the target point labeled. From these images a small number is chosen to create an initial point model by extracting the edge points from a region of interest around the annotated point. The contributions from the different images are fused and used as initial model. By employing the GHT and the discriminative training technique (DMC) with this model on a set of trainings images, first model point weights are learned.

After the first iteration, the estimated model is tested on the remainder of the development dataset. Images where the model performed poorly will be included into the set of training and model generation images, and another training iteration will be performed based on the new input. The training procedure stops if the localization error on the development dataset is below a certain threshold or if no further improvement is achieved.

If the model would only be generated from regions around the annotated point, as described above, there would be no chance to include information about confusable structures, since all given model points belong to the shape object and its surrounding. To incorporate anti-shape information into the model as well, a second region is extracted from the model images around the erroneous point that had the highest number of votes in the Hough space. The usage of these *anti-shape* points leads to a more discriminative model, which is able to distinguish between the object of interest and rivaling objects.

Since the processing time of the GHT depends strongly on the size of the model, we try to diminish the number of model points by keeping only relevant points after each iteration. To this end the points are sorted by their absolute weight and the number of points necessary to establish a minimal error on the training data is retained.

An overview of the complete procedure is given in Fig. 1.

2.4 Material and Task

The algorithm is applied to a large database of 740 long-leg radiographs from about 600 different patients. The field of view for most images includes the right and left leg from the ankle up to the hip, with varying artificial replacements of the hip and knee; some of the images cover only one leg or certain joints.

The images have an isotropic pixel spacing of 0.143 mm. Due to the high resolution and the resulting data size, the images were down-sampled for performance reasons to a pixel spacing of 1.14 mm using a Gaussian pyramid.

The given task is to localize all joints, namely femur, knee and ankle, for a subsequent segmentation of the bones as described in [4]. Annotations of the joints exist for all images and are utilized during training and as ground-truth to evaluate the localization accuracy. The annotations were performed several times by one observer, who obtained a mean intra-observer error of 2.3 mm for the femur, 1.3 mm for the knee and 2.6 mm in case of the ankle.

The models for this task are trained on right joints without pathological findings. From the complete database 80 images are randomly chosen as development dataset, such that between 50 and 60 images are available for each joint. The procedure starts on three images, from which one is used to generate an initial model. In each iteration the three images with the largest error are added to the training set. The image with the largest error is furthermore used to generate new model points. The training stops if an error of less than 5 Hough cells, which have a spacing of 2.29 mm, is obtained on each image of the dataset or if no further improvement is achieved by the training.

3 Results

The accuracy and localization rate achieved on a test dataset of 660 unseen images is stated in Table 1. As can be seen there, the localization performance is best for the knee, which is well visible in most images. The ankle obtains a larger error most likely due to the rotational freedom of the joint, which seems to be difficult to capture. The localization of the femur, yielding the highest error, is hampered by a low image contrast in that region. However, for many images with a wrong localization result, the second or third highest peak in the Hough space points to the correct cell, such that the target object could be localized in most images by keeping multiple candidates. Since the models were trained on images without pathological findings, we cannot expect them to be able to localize joints with artificial replacements or abnormalities. Nevertheless, good results are even achieved on these data (see Table 1).

For the further evaluation of the results, we will focus on the knee, which demonstrates best the advantages of the new procedure. The challenge for the knee is that the joint itself appears very similar for both legs such that surrounding structures, in particular the fibula bone, need to be included in the model to distinguish right from left. The evolution of the model for the right knee is shown in Table 2. In each iteration the number of training images is increased and new model points are appended until the model is capable of capturing the

Table 1. Results for the localization of the three joints on 660 unseen test images. The table states the size of the respective model, the mean error and the percentage of successfully localized joints.

	model size	w/o path. findings		art. replacements		abnormalities	
femur	1608	12.5 mm	74 %	14.6 mm	70 %	17.4 mm	50 %
knee	1923	4.3 mm	97 %	8.5 mm	85 %	6.8 mm	71 %
ankle	1187	9.8 mm	87 %	-	-	13.3 mm	66 %

Table 2. Evolution of the knee model. In the top part of the table the size of the trained model in each iteration is specified. The bottom part states the mean error in mm on the training and development images.

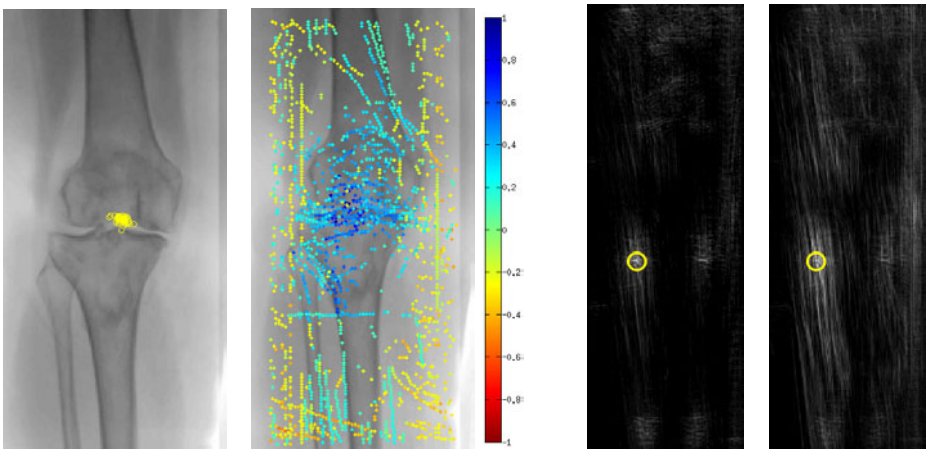
iteration	1	2	3	4
no. of model points	75	555	1817	1923
error on training data	1.71	2.06	3.26	1.57
error on development data	97.50	25.23	24.79	3.35

variability in size, shape and rotation of the target object. Using the model obtained from only one image in the first iteration, the wrong knee is chosen in 12 of 51 images. After four iterations the model is discriminative for the right knee and the joint is localized correctly in all images. The distribution of the localization error on unseen test data is visualized in Fig. 2(a).

Fig. 2(b) shows an image of the final weighted model for the knee, with color-coded model point weights. The plot reveals that the area between femur and tibia is the most discriminative for the knee, since these model points have the highest positive weights. Furthermore, the incorporation of variability in bone orientation of the femur and tibia in the model is visible. The size of the trained models is given in Table 1. Prospective research will aim at the further reduction of model size.

To reveal the impact of the model point weights on the GHT, we conducted some further experiments on the development dataset by localizing the knee with a standard unweighted GHT. In the first experiment, we generated an initial unweighted model from the four model images, which have been identified in the previous experiment. The mean error, obtained with this model, which should contain sufficient shape information for a correct localization, is 111.13 mm and the localization only succeeded on 23 of 53 images. This experiment reveals that the information about the importance of points is necessary and that sole contour information is not sufficient for a good localization.

In the second experiment, we used the final model of the knee obtained with the iterative approach, but ignored the weights in this case. The model points representing anti-shapes are excluded from the model, as well. Thereby, a mean localization error of 8.42 mm is obtained, which is a strong decline compared to



(a) (b) (c) (d)

Fig. 2. The left images show (a) the distribution of the error and (b) the weighted model overlaid on an image of the knee. Shape points are displayed in blue, anti-shape points in red. The right images show a comparison of Hough spaces obtained with (c) a weighted and (d) an unweighted model. In both cases the highest votes are accumulated close to the correct cell (yellow circle). However, the weighted model leads to a more focused Hough space. (For visualization purposes, the gray-value range is scaled to the value range of the Hough space in the left image.)

the mean error of 2.84 mm achieved by the weighted model. The impact of the weights on the GHT can also be seen in Fig. 2(c) and 2(d). Compared to the unweighted model without anti-shape information, the weighted model generates a more focused Hough space with a clear maximum in the correct cell.

4 Conclusion

A training approach to generate discriminative models for object localization with the GHT was presented. On a dataset of 660 unknown images, 74–97 % of the different joints were localized correctly with a mean error of 4.4–12.5 mm, which is remarkable considering the substantial anatomical variability of the data and the pixel size of 2.29 mm used in the GHT. The low resolution was chosen due to performance reasons. If a more precise localization is needed, the procedure could be embedded into a multi-level setting. The current results are used to initialize a segmentation procedure [4], which has a capture range of about 1 cm. The results of the segmentation can subsequently be used for orthopedic computations like the estimation of length, angles or density of bones.

The supervised training procedure runs fully automatic without the need of user interaction, and reliably results in discriminative models. By means of an iterative procedure, model points, which are discriminative and robust for the target object, are identified successively. Simultaneously, a small set of training

images is established, which represents the target object and its rivaling structures well. Manually determining this information, assuming it would be feasible, would be a highly demanding and time consuming task.

Further experiments have illustrated the importance of model point weights for a discriminative model. Especially negative weights play an important role since these anti-shape points aid in repelling the model from rivaling objects, thereby reducing false positive rates.

The proposed procedure is applicable to any localization task where the target object is well defined by its shape and has proven to significantly improve localization accuracy compared to a standard unweighted GHT.

Acknowledgments. The authors would like to thank the Dartmouth-Hitchcock Medical Center and Diagnostic X-Ray, Philips Healthcare for providing the radiographs used in this study. This work is partly funded by the Innovation Foundation Schleswig-Holstein under the grant 2008-40 H.

References

1. Ballard, D.H.: Generalizing the Hough Transform to Detect Arbitrary Shapes. Pattern Recognit. 13(2), 111–122 (1981)
2. Beyerlein, P.: Discriminative Model Combination. In: Proceedings of the 1998 IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 481–484 (1998)
3. Deselaers, T., Keysers, D., Ney, H.: Improving a Discriminative Approach to Object Recognition using Image Patches. In: Kropatsch, W.G., Sablatnig, R., Hanbury, A. (eds.) DAGM 2005. LNCS, vol. 3663, pp. 326–333. Springer, Heidelberg (2005)
4. Gooßen, A., Hermann, E., Gernoth, T., Pralow, T., Grigat, R.-R.: Model-Based Lower Limb Segmentation Using Weighted Multiple Candidates. In: Bildverarbeitung für die Medizin, pp. 276–280. Springer, Heidelberg (2010)
5. Heimann, T., van Ginneken, B., Styner, M.A., Arzhaeva, Y., Aurich, V., Bauer, C., Beck, A., Becker, C., Beichel, R., Bekes, G., Bello, F., Binnig, G., Bischof, H., Bornik, A., Cashman, P., Ying, C., Cordova, A., Dawant, B.M., Fidrich, M., Furst, J.D., Furukawa, D., Grenacher, L., Hornegger, J., Kainmuller, D., Kitney, R.I., Kobatake, H., Lamecker, H., Lange, T., Lee, J., Lennon, B., Li, R., Li, S., Meinzer, H.-P., Nemeth, G., Raicu, D.S., Rau, A.-M., van Rikxoort, E.M., Rousson, M., Rusko, L., Saddi, K.A., Schmidt, G., Seghers, D., Shimizu, A., Slagmolen, P., Sorantin, E., Soza, G., Susomboon, R., Waite, J.M., Wimmer, A., Wolf, I.: Comparison and Evaluation of Methods for Liver Segmentation from CT Datasets. IEEE Trans. Med. Imaging 28(8), 1251–1265 (2009)
6. Heimann, T., Münziger, S., Meinzer, H.-P., Wolf, I.: A Shape-Guided Deformable Model with Evolutionary Algorithm Initialization for 3D Soft Tissue Segmentation. In: Information Processing in Medical Imaging, pp. 1–12 (2007)
7. Jaynes, E.T.: Information Theory and Statistical Mechanics. Phys. Rev. 106(4), 620–630 (1957)
8. Lee, Y., Hara, T., Fujita, H., Itoh, S., Ishigaki, T.: Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique. IEEE Trans. Med. Imaging 20(7), 595–604 (2001)

9. Leibe, B., Leonardis, A., Schiele, B.: Robust Object Detection with Interleaved Categorization and Segmentation. Int. J. Computer Vis. 77(1-3), 259–289 (2008)
10. Linguraru, M.G., Vercauteren, T., Reyes-Aguirre, M., González Ballester, M.A., Ayache, N.: Segmentation Propagation from Deformable Atlases for Brain Mapping and Analysis. Brain Research Journal 1(4) (2007)
11. Maji, S., Malik, J.: Object Detection using a Max-Margin Hough Transform. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1038–1045 (2009)
12. Martín Recuero, A.B., Beyerlein, P., Schramm, H.: Discriminative Optimization of 3D Shape Models for the Generalized Hough Transform. In: 7th International Conference and Workshop on Ambient Intelligence and Embedded Systems (2008)
13. Ruppertshofen, H., Lorenz, C., Beyerlein, P., Salah, Z., Rose, G., Schramm, H.: Fully Automatic Model Creation for Object Localization utilizing the Generalized Hough Transform. In: Bildverarbeitung für die Medizin, pp. 281–285. Springer, Heidelberg (2010)
14. Schramm, H., Ecabert, O., Peters, J., Philomin, V., Weese, J.: Towards Fully Automatic Object Detection and Segmentation. In: SPIE Medical Imaging, p. 614402 (2006)
15. Seghers, D., Slagmolen, P., Lambelin, Y., Hermans, J., Loeckx, D., Maes, F., Suetens, P.: Landmark based liver segmentation using local shape and local intensity models. In: MICCAI Workshop on 3D Segmentation in the Clinic: a Grand Challenge, pp. 135–142 (2007)
16. Zheng, Y., Georgescu, B., Comaniciu, D.: Marginal Space Learning for Efficient Detection of 2D/3D Anatomical Structures in Medical Images. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 411–422. Springer, Heidelberg (2009)

Topology Noise Removal for Curve and Surface Evolution

Chao Chen[1,⋆] and Daniel Freedman[2]

[1] IST Austria (Institute of Science and Technology Austria)
Vienna University of Technology, Austria
chao.chen@ist.ac.at
[2] Hewlett-Packard Laboratories, Israel
daniel.freedman@hp.com

Abstract. In cortex surface segmentation, the extracted surface is required to have a particular topology, namely, a two-sphere. We present a new method for removing topology noise of a curve or surface within the level set framework, and thus produce a cortical surface with correct topology. We define a new energy term which quantifies topology noise. We then show how to minimize this term by computing its functional derivative with respect to the level set function. This method differs from existing methods in that it is inherently continuous and not digital; and in the way that our energy directly relates to the topology of the underlying curve or surface, versus existing knot-based measures which are related in a more indirect fashion. The proposed flow is validated empirically.

1 Introduction

Active contour model, first introduced by Kass *et al.* [11], is a well-known tool to perform the task of image segmentation, namely, computing a *contour* which separates the image (the *domain of interest*) into two parts, inside and outside. In such model, one evolves an initial contour according to the prescribed differential equations, until a steady-state is reached. The steady-state is then taken to be the desired contour.

In the level set framework, the contour is implicitly represented as the zero level set of a *level set function*, $\phi : \Omega \to \mathbb{R}$, where Ω is the domain of interest. ϕ is often taken as the *signed distance function*, whose absolute value is the distance to the contour, and whose sign is negative for points inside the contour and positive for points outside the contour[1]. Instead of evolving the contour, one then evolves ϕ and takes the zero level set of its steady-state as the final contour. In this paper, we take ϕ as the signed distance function for ease of illustration. However, our method could be applied to any level set function.

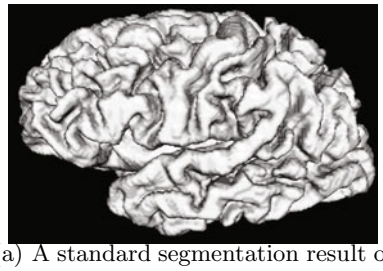
⋆ Partially supported by the Austrian Science Fund under grant P20134-N13.

[1] In some instances in the literature, the sign convention is the opposite of what we have described.

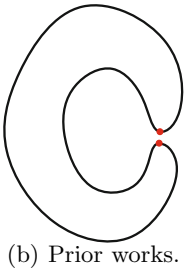
B. Menze et al. (Eds.): MICCAI 2010 Workshop MCV, LNCS 6533, pp. 31–42, 2011.

The level set method brings various advantages: parameterization free, extensible to any dimension, numerically stable. More importantly, it can handle changes in contour topology for free, due to the implicit representation.

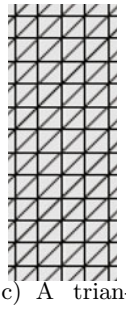
However, in certain applications, the topological flexibility is not desirable. The framework cannot distinguish meaningful topological features from noise, even if we have prior knowledge of the contour's topology. For example, in segmentation of a human cortex (Figure 1(a)), it would be beneficial if the extracted surface is homeomorphic to a two-sphere [10]. However, due to the complex geometry of a cortex surface, standard level set method (i.e. geodesic active contour [3,12], Chan-Vese [4]) would not achieve the correct topology. The narrow seam between the two hemispheres is especially challenging. In Figure 4(e), we show part of a standard segmentation result, namely, the part between the two hemispheres. Many holes appear, corresponding to small handles between the two hemispheres.



(a) A standard segmentation result of a human cortex, with topology noise.



(b) Prior works.



(c) A triangulation

Fig. 1.

Prior Works. Han *et al.* [10] proposed a digital algorithm which prevent topology changes during the evolution. Whenever a pixel's level set function value is updated and the sign of the function value changes, a check is applied to this pixel to make sure that it does not change the (digital) topology of the contour. If it does lead to a topological change, the pixel's function value is only updated to almost zero yet with the same sign as its old value. However, this topology control is detached from the energy minimization which drives the rest of the level set evolution. As a result, the method leads to undesirable artifacts, such as a contour which is a single pixel away from the wrong topology. Various digital topology methods have been developed, to achieve different topology constraints [1,14].

In order to produce more natural results, other methods [16,9] incorporate topology control through the introduction of an extra term in the energy functional to be minimized. The energy minimization framework then naturally encourages the contour to have the correct topology. These methods use a common intuition, based on knot theory: penalize the energy when two different parts of the contour meet (the two red points in Figure 1(b)).

All these methods prevent topology changes and maintain a same topology through the evolution. However, they cannot fix incorrect topology that already exists. This topological rigidity is inconsistent with the level set framework (Please see the end of this section for more discussion). In addition, the energy-based methods only prevent topology changes which arise from the merging of components. However, there are other types of topology change which may occur, such as the creation of new components which are far away from other components of the contour.

Contributions. We propose a new method for topology control within the level set framework. In this paper, the goal is to ensure that the contour (d-manifold) is homeomorphic to a d-sphere; however, the method can be extended to more general topological priors. For the remainder of this paper, "correct topology" will mean the topology of the d-sphere. We have two major contributions.

1. A measure of the contour's topology noise. Based on a recent theoretical result in computational topology [7], we define the *total robustness* of a contour, which, in a sensible way, indicates how close a contour is to having the correct topology. This measure is the sum of the robustness of individual topology features, where the robustness of a class measures how easy it is to "shake off" this feature by perturbing the contour.

2. A flow which drives the contour to the correct topology. Using the concept of total robustness, we compute a flow which pushes the level set function and thus the contour towards the correct topology. Specifically, we compute the functional derivative of the total robustness with respect to ϕ, and evolve ϕ according to gradient descent. On its own, this flow leads to a global minimum of the total robustness, and thus a contour with correct topology. Note that in practice, this flow will be combined with other flows, such as those designed for image segmentation; in this scenario, the global minimum property may be lost.

Unlike previous works, our method allows flexibility in the topology of the contour *during the evolution*, while ensuring that the topology of the *final* contour is correct. This is beneficial for two reasons: (1) the initial contour could have a different topology than the desired output contour (i.e. results from other segmentation model); (2) we do not need to worry about potential topology changes due to the discretization of the time step. Furthermore, compared with energy-based methods, our method addresses all types of "incorrect" topology, not just those which arise from merging of components as in [16,9].

Our flow changes the evolving level set function as locally as possible, and thus, will not change the geometry of the contour much. This flow could be used to post-process a standard segmentation result, namely, correct the topology of the result. In such scenario, it plays a similar role to topology correcting methods [15,17], which locate topology defects of a given surface and correct them locally. Unlike these methods, our method retains the level set framework due to the advantages described above (numerical stability, lack of need for parameterization, etc.), as well as the fact that it is the "native language" for many applications in vision and graphics. Furthermore, our flow could be easily

combined with other flows (i.e. geodesic active contour), so that the correction result is natural (See Section 4).

In this paper, for ease of exposition, we focus our discussion on the case when the contour C is a one-manifold and the domain is a 2D image ($\Omega \subset \mathbb{R}^2$). However, our algorithm can be naturally extended to d-manifolds in $(d+1)$-dimensional domain. In specific, we implemented and experimentally verified our method on 3D images.

2 Background

Topology Features. In this paper, instead of the topology of the contour, C, we focus on the topology of the object enclosed by the contour, $O = \phi^{-1}(-\infty, 0] = \{x \in \Omega \mid \phi(x) \leq 0\}$. It is provable that if two objects have different topology, then the corresponding contours have different topology[2]. The "correct topology" constraint, namely, C is homeomorphic to a d-sphere, is equivalent to the constraint that O is homeomorphic to a $(d+1)$-dimensional ball.

The topology features we discuss are homology classes over $\mathbb{Z}_2$ field, which are well studied in algebraic topology (see [13] for a formal introduction). In 2D images, 0D and 1D classes of an object are the components and the holes, respectively. For example, in Figure 3(a), the object has four components and three holes. In 3D image, 0D, 1D and 2D classes are the components, handles, and voids of the object. Formally, the modulo-2 sum of any set of homology classes is also a homology class. The group of classes form a vector space. In this paper, however, we only focus on a canonical basis of such vector space, and call this basis the set of topology features. For example, in Figure 3(a), we only consider the set of four components and the set of three holes, denoted as $\mathsf{H}_0(O)$ and $\mathsf{H}_1(O)$ respectively. Their union is denoted as $\mathsf{H}(O)$ for convenience.

Critical Points. Given a domain Ω and a function $\phi : \Omega \to \mathbb{R}$, as we increase a threshold t from $-\infty$ to $+\infty$, the sublevel set, $\phi^{-1}(-\infty, t]$, grows from the empty set to the entire domain. During this process, topology of the sublevel set come into existence ("birth") and disappear ("death"). The points in the domain at which such topology changes happen are called *critical points*. Their function values are *critical values*. This notation, defined by Cohen-Steiner *et al.* [5], is an extension of Morse theory. Please note that we assume zero is a *regular value*, namely, a value which is not critical. In other words, we assume no topology change happens at $t = 0$.

In Figure 2(Left), on the top, we show a contour as well as the object enclosed, which has 3 components. Below it, we draw the graph of the corresponding signed distance function ϕ (using ϕ as the z coordinate). For ease of understanding, we also draw the contour in the graph, which is the intersection of the graph and the $\{z = 0\}$ plane. There are seven critical points, highlighted with red marks.

[2] The converse, however, is not true. For example, given a contour homeomorphic to the disjoint union of two one-spheres, the object enclosed could be either an annulus or the disjoint union of two disks.

Four *minimal* points, $m_0, \cdots, m_3$, are the birth places of four components, born at $\phi(m_0), \cdots, \phi(m_3)$ respectively. Three *saddles* s_1, s_2, s_3 are the the points at which components are merged, corresponding to the death of 3 components. There are also *maximal* points, at which holes disappear during the growth (there are three maximal points in Figure 3(a)).

In application, we compute these critical points and their characteristic (minimal, saddle, or maximal) in a discretized framework. We triangulate the image, Ω, using the set of pixels as vertices (Figure 1(c)). We approximate the growth of the sublevel set, $\phi^{-1}(-\infty, t]$, by a growing subcomplex, namely, a subset of simplices (vertices, edges and triangles). Each simplex, σ, has a distinct function value and enters the growing subcomplex as soon as the threshold $t \geq \phi(\sigma)$. Critical points then correspond to simplices which change the topology of the growing subcomplex. This method, known as the *persistence homology algorithm* [6], is the foundation of the algorithm to compute robustness, which will be used later.

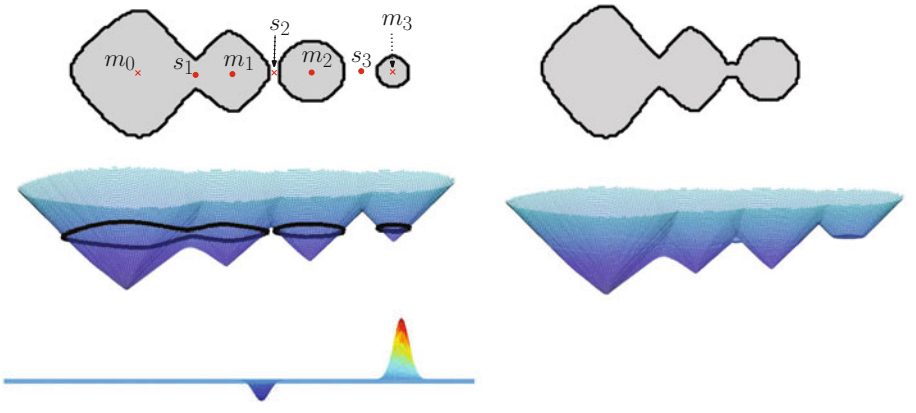


Fig. 2. Left: a contour with topology noise, the graph of its signed distance function, ϕ, and the graph of the flow, $\partial\phi/\partial t = -\delta E/\delta\phi$.
Right: a perturbation of ϕ and its contour. Topology noise is removed.

Robustness. Edelsbrunner *et al.* [7] defined a measure of homology classes of the object O, $\rho_\phi : \mathsf{H}(O) \to \mathbb{R}$. Intuitively, given a homology class $\alpha \in \mathsf{H}(O)$, its *robustness*, $\rho_\phi(\alpha)$ measures how easily it is to kill α by perturbing the function. In other words, $\rho_\phi(\alpha)$ is the minimal error we can tolerate to get a function which approximates ϕ and kills α. Different classes of an object have different robustness. The ones with small robustness are considered noise, and will be removed by our method.

Given $r \geq 0$, we call $h : \Omega \to \mathbb{R}$ an *r-perturbation* of ϕ if the L_∞ distance between the two functions is upperbounded by r, formally, $\|h - \phi\|_\infty = \max_{x \in \Omega} |h(x) - \phi(x)| \leq r$. Define the robustness, $\rho_\phi(\alpha)$, as the minimal r such that there exists an r-perturbation of ϕ, h, such that α is dead in the perturbed

object, $h^{-1}(\infty, 0]$. By dead, we mean that α either disappears or is merged with some other class which is more robust.

For example, in Figure 2(Left), to kill the middle component of O, we have two options: (1) Increase function values of points near the minimal point m_2. In the function graph, this is equal to pushing the cone tip up so that $h(m_2) = 0 + \epsilon$, ϵ positive and arbitrarily close to zero. The component would then disappear in $h^{-1}(-\infty, 0]$. (2) Decrease function values of points near the saddle point s_2. In the function graph, we drag the saddle down so that $h(s_2) = 0 - \epsilon$. The component would then merge with the left component, which is more robust.

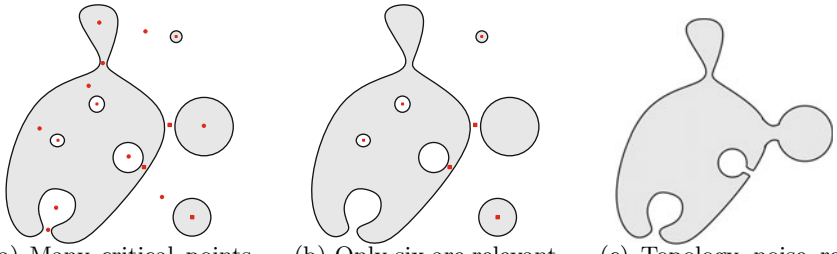
The two options lead to two new functions, h_m and h_s, which are $|\phi(m_2)|$- and $|\phi(s_2)|$-perturbations of ϕ, respectively. Since here $|\phi(m_2)| > |\phi(s_2)|$, we choose the the second option as the best perturbation, whose L_∞ distance from ϕ, $|\phi(s_2)|$, is thus the robustness of the corresponding component. Similarly, we determine that the robustness of the right component and the left component are $|\phi(m_3)|$ and $|\phi(m_0)|$ respectively (note that $|\phi(m_3)| < |\phi(s_3)|$)[3]. In Figure 2(Right), we show a $|\phi(m_3)|$-perturbation which kills the right and the middle component, but not the left one.

It is noticeable that the robustness has a close relationship with critical values and critical points. In fact, each class $\alpha \in \mathsf{H}(O)$ corresponds to a specific critical point, c_α. The absolute value of the corresponding critical value is the robustness, namely, $\rho_\phi(\alpha) = |\phi(c_\alpha)|$. Bendich *et al.* [2] provided an *robustness algorithm* to compute the robustness of each homology class of the object, as well as the associated critical point. The algorithm is based on the persistent homology algorithm, and takes cubic time in the worst case and almost linear time in practice.

In general, given a function ϕ, there are a lot of critical points, each is relevant to the topology of some sublevel set. Using robustness algorithm, we could identify those which are the most relevant to the topology of the object O. Changing their neighborhood would kill topology noise. And furthermore, this changing is local and thus will not harm the geometry of the object much. In Figure 3(a), we show a contour and its object, with topology noise. The level set function ϕ has many critical points (red marks). The robustness algorithm finds the 6 critical points corresponding to the 6 undesired homology classes of O (Figure 3(b)). Changing their neighborhood leads to a perturbation of ϕ, and thus a new object with no topology noise (Figure 3(c)).

Although the algorithm could identify relevant critical points, it is not clear how to change the local neighborhood of them to get the correct topology while keeping the geometry of O as intact as possible. In this paper, we define an energy term defined on ϕ. This energy term is minimized if and only if the contour/object has the correct topology. We then use gradient descent to iteratively evolve the level set function until the contour/object has the correct topology. This gradient only changes a small neighborhood of these relevant critical points and thus ensure the geometry is less influenced.

[3] In the rare case that $|\phi(m_i)| = |\phi(s_i)|$, there is ambiguity. See Appendix A for how to deal with such cases.



(a) Many critical points (red marks). (b) Only six are relevant to topology noise of O. (c) Topology noise removed.

Fig. 3. A contour with topology noise

3 Method

The Energy Term: Total Robustness. Given a contour C, and the level set function ϕ, we define an energy term as the sum of the k-th power of robustness of topology noise. Formally, the *degree-k total robustness* of O is

$$Rob_k(O) = \sum_{\alpha \in \mathsf{H}(O)} \rho_\phi(\alpha)^k - \left(\max_{\alpha \in \mathsf{H}(O)} \rho_\phi(\alpha) \right)^k . \tag{1}$$

This energy sums over all classes of O except for the most robust one, which is the component we want to keep. This component is born at the global minimal point, c_{min}. Its robustness is $|\phi(c_{min})|$. Thus the last term of Equation (1) is $-|\phi(c_{min})|^k$.

Assuming O has at least one component, the total robustness is non-negative. It reaches its global minimum, zero, if and only if the contour/object has the correct topology.

We are now ready to compute the flow which drives the contour C, and its signed distance function ϕ, towards the minimum of the total robustness. Denoting the energy $E(\phi) = Rob_k(\phi)$, the goal is to compute the functional derivative $\delta E/\delta\phi$, and thus the desired flow $\partial\phi/\partial t = -\delta E/\delta\phi$. By gradient descent, the flow minimizes E.

Basing on the robustness theory, we could prove that the difference between two level set functions upperbounds the difference between their total robustness. Therefore, the energy term E is continuous. Unfortunately, E is not differentiable everywhere. We will now compute the derivative for cases when E is differentiable, and explain the intuition. The non-differentiable cases will be discussed in Appendix A. We prove that the set of functions at which E is non-differentiable is measure zero in the space of functions (Theorem 1). This theorem leads to the construction of a smoothed approximation of E, whose corresponding flow can be efficiently computed.

For the rest of the section, we use degree-3 total robustness as the energy, $E(\phi) = Rob_3(\phi)$.

The Functional Derivative. We first assume that all critical points have distinct values. In this case, E is differentiable. Recall that c_α is the critical point of ϕ associated to a homology class $\alpha \in \mathsf{H}(O)$, such that $|\phi(c_\alpha)| = \rho_\phi(\alpha)$. We can rewrite E as a function depending on ϕ and relevant critical points,

$$E(\phi) = \sum_{\alpha \in \mathsf{H}(O), c_\alpha \neq c_{min}} sign\left(\phi(c_\alpha)\right) \phi^3(c_\alpha)$$

The set of relevant critical points $\{c_\alpha | \alpha \in \mathsf{H}(O)\}$ depends on the function ϕ and can be computed by running the robustness algorithm once. Recall that the algorithm discretizes the image into a triangulation and approximate the growing sublevel set by a growing subcomplex. In this case, a sufficiently small perturbation of ϕ will preserve the order in which simplices entering the growing subcomplex. This would lead to the fact that the output of the algorithm remains almost the same. In specific, the critical point associated to each homology class remains the same. The sign of its critical value remains the same. Only its critical *value* will change (according to the perturbation).

Finally, we can rewrite $E(\phi)$ as a functional and compute the functional derivative.

$$E(\phi) = \int_\Omega \left(\sum_{\alpha \in \mathsf{H}(O), c_\alpha \neq c_{min}} \delta(x - c_\alpha) \right) sign\left(\phi(x)\right) \phi^3(x) dx$$

$$\frac{\delta E}{\delta \phi} = 3 \left(\sum_{\alpha \in \mathsf{H}(O), c_\alpha \neq c_{min}} \delta(x - c_\alpha) \right) sign\left(\phi(x)\right) \phi^2(x) \tag{2}$$

where $\delta(x - c_\alpha)$ is a Dirac delta function. The second equation is due to the Euler-Lagrange equation and the fact that c_α's and $sign(\phi(c_\alpha))$'s are constant for a small perturbation of ϕ. In our implementation, we use a standard trick from the level set literature and relax $\delta(x - c_i)$ to a bounded C^1 approximation with width σ, $\delta_\sigma(x - c_i)$, e.g. a Gaussian with variance σ^2.

Next, we illustrate the intuition of the functional derivative in Equation (2). The derivative $\delta E/\delta\phi$ is the weighted sum of a set of Dirac delta functions (or Gaussians in the relaxed case), centered at relevant critical points, except for c_{min}. The weights are proportional to the square of the corresponding critical values multiplied by its sign. During the evolution, we update ϕ according to $\partial\phi/\partial t = -\delta E/\delta\phi$. This pushes function values near the relevant critical point towards zero, and thus shrinks the corresponding robustness. In Figure 2(Left), we draw the graph of the flow, $\partial\phi/\partial t$, at the bottom. The flow lifts the minimal point, m_3, corresponding to the right component. For the middle component, the flow suppresses the saddle point, s_2. The rate of lifting or suppression is proportional to the square of the corresponding robustness, so that classes with large robustness have their robustness decreased faster than those with small robustness.

Although the energy is not everywhere differentiable, we could construct a smoothed approximation E_λ. The functional derivative $\delta E_\lambda/\delta\phi$ is the same as

Equation (2) at ϕ further away from the non-differentiable set, and can be computed by Monte Carlo method otherwise. Please see Appendix A for details.

Global Minimum. Typically gradient descent leads only to a local minimum of the energy. In our case, we in fact achieve the global minimum, i.e. zero energy. The reason is that Equation (2) is zero if and only if one class left, when $E(\phi) = 0$. Therefore, the output of our flow has the "correct topology".

4 Implementation and Experimental Results

For implementation of the level set method, we use the freely available implementation by Zhang [18]. We compute robustness using our own implementation of the robustness algorithm [2]. The flow is then computed according to Equation (2).

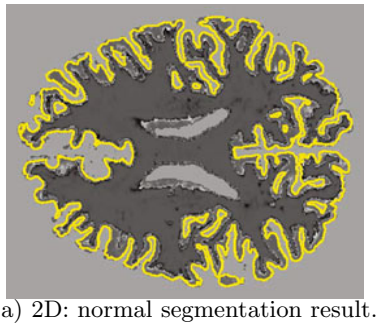
We use our method as a post-processing step. We first apply a standard segmentation algorithm (GAC or Chan-Vese). Using this output as an initial contour, we evolve with our flow to correct the topology. Please note that this scheme will not fit other topological control methods, as they require that the topology be held constant throughout the evolution.

We verify the method on medical images. In Figure 4(a) and 4(b), we segment the image of a slice of brain white matter. Our method is able to correct various types of topological noise, and recovers a contour homeomorphic to a one-sphere. Please note that these images are only a proof of concept. The boundary of a slice of brain white matter is not necessarily homeomorphic to a one-sphere.

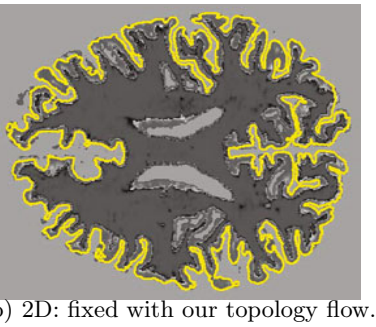
In a 3D MR image, however, a cortex surface would be homeomorphic to a two-sphere. We verify our method with a 3D MR image (from the example data of FreeSurfer [8]). We compare the initial surface (the result of a standard segmentation) and the surface after applying our flow. In Figure 4(c) and 4(d), we show a handle that is removed after applying our flow. In Figure 4(e) and 4(f), we show the part of the cortex surface between the left and right hemispheres. Note the holes in the surface corresponds to handles between the two hemispheres. Our flow will fix the topology by either remove small handles, or merging handles nearby. Therefore, after applying the flow, the surface has only one big hole left.

We note the choice of the width of the Gaussian kernel, σ, is essential in the converging speed. The bigger σ is, the faster the result converges. On the other hand, σ decides how natural the fixing result could be. In our experiments, we use a Gaussian kernel with $\sigma = 0.8$. Further study of this parameter will be done in the future.

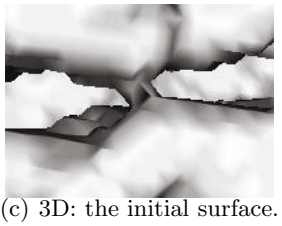
To improve the speed and decrease the memory consumption, we choose to discretize 3D images with cubical complexes. Such complex consists of cubes, corresponding to voxels, as well as their facets, including faces, edges and vertices. We approximate the growth of the sub-level set $\phi^{-1}(-\infty, t]$ by a growing subcomplex, consisting of cubes (and their facets) whose function value is no greater than t. To compute topology, we choose different connectivities (preferably 18 or 26) of the subcomplex. This would natural leads to an algorithm to compute critical points and values, and thus robustness.



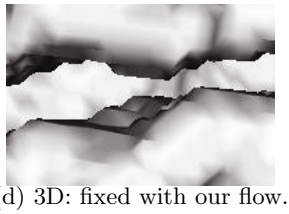
(a) 2D: normal segmentation result.



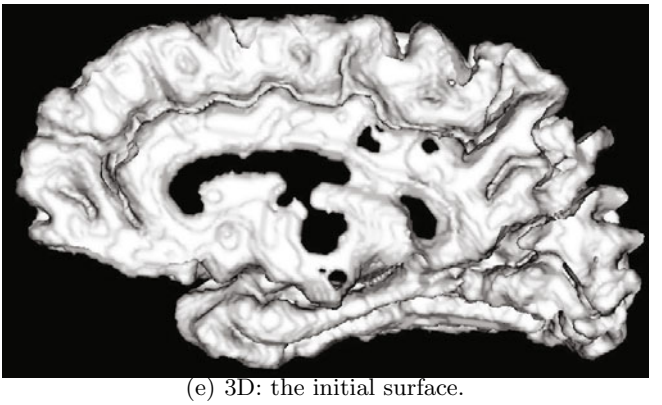
(b) 2D: fixed with our topology flow.



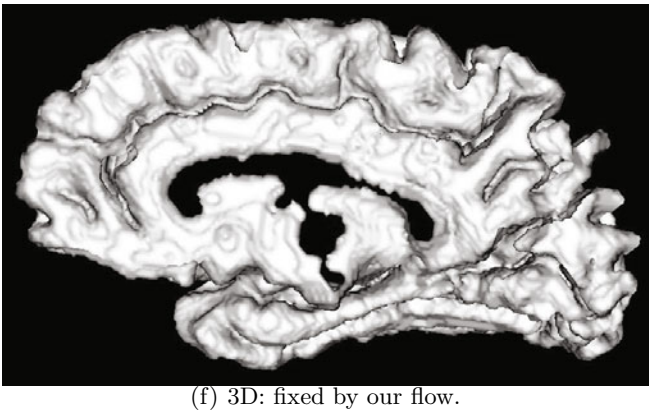
(c) 3D: the initial surface.



(d) 3D: fixed with our flow.



(e) 3D: the initial surface.



(f) 3D: fixed by our flow.

Fig. 4.

We run experiments on a computer with 10 2.53GHz Xeon Processors and 96GB RAM. For a 2D image with 200×200 pixels, the evolution takes 10 iterations. Each iteration costs 1 to 2 minutes. For a 3D image with 256×256×256 voxels, the evolution takes 10 to 20 iterations. Each iteration costs 6 to 7 minutes. One bottleneck of our method is the memory cost. For a 200×200 2D image, the memory cost is 500MB. For a 256×256×256 3D image, the memory cost is 7GB. This has close relationship with the huge number of simplices in a triangulation. For example, for a 256×256×256 3D image, the number of simplices is 26×16.5 million.

Acknowledgement

We thank Helena Molina-Abril for very helpful discussion. We thank anonymous reviewers for helpful comments.

References

1. Bazin, P.-L., Pham, D.L.: Topology correction of segmented medical images using a fast marching algorithm. Computer Methods and Programs in Biomedicine 88(2), 182–190 (2007)
2. Bendich, P., Edelsbrunner, H., Morozov, D., Patel, A.: The robustness of level sets. In: de Berg, M., Meyer, U. (eds.) ESA 2010. LNCS, vol. 6346, pp. 1–10. Springer, Heidelberg (2010)
3. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. International Journal of Computer Vision 22(1), 61–79 (1997)
4. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Transactions on Image Processing 10(2), 266–277 (2001)
5. Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Stability of persistence diagrams. Discrete & Computational Geometry 37, 103–120 (2007)
6. Edelsbrunner, H., Harer, J.: Computational Topology. An Introduction. Amer. Math. Soc., Providence (2009)
7. Edelsbrunner, H., Morozov, D., Patel, A.: Quantifying transversality by measuring the robustness of intersections (2009) (preprint)
8. Freesurfer, `http://surfer.nmr.mgh.harvard.edu/`
9. Guyader, C.L., Vese, L.A.: Self-repelling snakes for topology-preserving segmentation models. IEEE Transactions on Image Processing 17(5), 767–779 (2008)
10. Han, X., Xu, C., Prince, J.L.: A topology preserving level set method for geometric deformable models. IEEE Trans. Pattern Anal. Mach. Intell. 25(6), 755–768 (2003)
11. Kass, M., Witkin, A.P., Terzopoulos, D.: Snakes: Active contour models. International Journal of Computer Vision 1(4), 321–331 (1988)
12. Kichenassamy, S., Kumar, A., Olver, P.J., Tannenbaum, A., Yezzi, A.J.: Gradient flows and geometric active contour models. In: ICCV, pp. 810–815 (1995)
13. Munkres, J.R.: Elements of Algebraic Topology. Addison-Wesley, Redwook City (1984)
14. Ségonne, F.: Active contours under topology control - genus preserving level sets. International Journal of Computer Vision 79(2), 107–117 (2008)

15. Ségonne, F., Pacheco, J., Fischl, B.: Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. IEEE Trans. Med. Imaging 26(4), 518–529 (2007)
16. Sundaramoorthi, G., Yezzi, A.J.: Global regularizing flows with topology preservation for active contours and polygons. IEEE Transactions on Image Processing 16(3), 803–812 (2007)
17. Yotter, R.A., Dahnke, R., Gaser, C.: Topological correction of brain surface meshes using spherical harmonics. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 125–132. Springer, Heidelberg (2009)
18. Zhang, Y.: The matlab toolbox for 2d/3d image segmentation using level-set based active contour/surface with aos scheme, `http://ecson.org/resources/active_contour_segmentation.html`

A Non-differentiable Cases

In this section, we discuss the non-differentiable cases of the energy E.

Previously, we assumed that all critical points have distinct values. If we relax such constraint, there are cases at which E is non-differentiable. For example, in Figure 2(Left), if we let $|\phi(s_3)| = |\phi(m_3)|$, the critical point associated to the right component could be either m_3 or s_3. Thus the functional derivative is not well defined. In general, non-differentiable cases happen when critical points share a same value, causing ambiguity of associating c_α to each class α.

On the other hand, we have the following theorem (The proof is omitted).

Theorem 1. *The set of points at which $E(\phi)$ is non-differentiable, denoted as ND, is measure 0.*

This theorem means that E is differentiable almost everywhere, leading to the possibility of approximating E with a smoothed version, E_λ, whose derivative can be efficiently computed.

Denote Φ as the space of functions ϕ, we define the smoothed energy term as

$$E_\lambda(\phi) = \int_\Phi E(\xi) G_\lambda(\phi - \xi) d\xi$$

where $G_\lambda : \Phi \to \mathbb{R}$ is a C^1 function defined on function space with a compact support $B_\lambda(0) = \{\xi \in \Phi \mid \|\xi\|_2 \leq \lambda\}$, and which integrates to 1[4].

Due to Theorem 1 and a change of variables, the derivative is

$$\frac{\delta E_\lambda}{\delta \phi} = \int_{B_\lambda(\phi) - ND} \frac{\delta E}{\delta \phi}(\phi - \xi) G_\lambda(\xi) d\xi$$

For sufficiently small λ, the λ-ball $B_\lambda(\phi)$ around most points ϕ will not intersect the set of non-differentiable functions; as a result, $\partial E_\lambda / \partial \phi \approx \partial E / \partial \phi$, where the latter can be evaluated directly from Equation (2) without an integral. In cases where $B_\lambda(\phi)$ does indeed intersect the set of non-differentiable functions, the integral may be approximated by Monte Carlo methods.

[4] This is analogous to smoothing a piecewise smooth function defined on Euclidean space by the convolution with a smooth Gaussian kernel.

Exploring Cortical Folding Pattern Variability Using Local Image Features

Rishi Rajalingham[1], Matthew Toews[2], D. Louis Collins[3], and Tal Arbel[1]

[1] Center for Intelligent Machines, McGill University
[2] Brigham and Women's Hospital, Harvard Medical School
[3] Montreal Neurological Institute, McGill University

Abstract. The variability in cortical morphology across subjects makes it difficult to develop a general atlas of cortical sulci. In this paper, we present a data-driven technique for automatically learning cortical folding patterns from MR brain images. A local image feature-based model is learned using machine learning techniques, to describe brain images as a collection of independent, co-occurring, distinct, localized image features which may not be present in all subjects. The choice of feature type (SIFT, KLT, Harris-affine) is explored with regards to identifying cortical folding patterns while also uncovering their group-related variability across subjects. The model is built on lateral volume renderings from the ICBM dataset, and applied to hemisphere classification in order to identify patterns of lateralization based on each feature type.

1 Introduction

The cerebral cortex is characterized by complex folding patterns created by ridges called gyri, and fissures called sulci. The exploration and identification of these cortical folding patterns is important in the development of a general functional atlas of the cortex, as folds are linked to the division of functional areas in the brain. However, this task is difficult due to the fact that cortical structures may vary significantly in shape, size and appearance across subjects, or may not even be present in all subjects. This phenomenon is known as inter-subject variability, the variation in feature morphology across different subjects caused by inherent anatomic variability or pathology, and renders it difficult even for experts to identify folding patterns [1].

Uncovering cortical variability can be instrumental in the understanding of how various anatomical structures express themselves within and across subject groups; an appropriate quantification of variability can answer fundamental questions in medicine, such as which cortical folds are common or rare across a particular population. Quantifying anatomical variability at the group level, based for example on pathology, can give insight as to how similarities or differences in expression of particular folds are related to subsets of the population. This can be then be used as a means of evaluating similarity of a new subject to established groups.

B. Menze et al. (Eds.): MICCAI 2010 Workshop MCV, LNCS 6533, pp. 43–53, 2011.

The conventional approach to examine and identify cortical folding patterns relies on manual labelling by experts, which is a tedious task that is subject to inter-rater variability. For this reason, a significant amount of work has been done on automatic, data-driven techniques to learn cortical folding patterns and study their variability using morphometric techniques on magnetic resonance imaging (MRI) data of the brain. Morphometry is the analysis of variation in form (appearance, shape or size) of objects. In this context, it aims to identify anatomical variability of cortical folds between subjects. Most automatic analyses rely firstly on warping each brain image to a common reference frame, using voxel intensities [2], landmark objects [3], cortex-specific features [4], to name a few, and computing statistical quantities based on this registration. These methods assume or even force a one-to-one correspondence between subject brains that may not generally exist. Anatomical variability in folding patterns can lead to imperfect registration, resulting in the averaging out of structural differences. Consider the case where a particular anatomical structure expresses itself in multiple, distinct types across a population, or only expresses itself in part of the population. This type of group-related variability cannot be uncovered by most morphometric techniques.

A number of methods address this issue by loosening the one-to-one correspondence assumption, by means of multiple atlases [5] or atlas stratification [6], or residual error components [7]. Feature based morphometry (FBM) is a probabilistic, parts-based model of images across subjects which effectively ignores the lack of a one-to-one correspondence in inter-subject registration, and rather attempts to identify local regions of images whose occurrences in groups is statistically significant [9]. This method has been applied to 2D MR slices to model anatomical variability [8], and volumetric brain images to discover group-related anatomical patterns [9]. It has had some success with surface renderings of the cortex [10], but was limited by the choice of local feature representation.

The main contribution of this paper is the introduction of an FBM technique for automatically learning new, unlabelled cortical folding patterns from a large set of subject brain images, attempting not only to cope with variability, but to uncover it in a completely data-driven, bottom-up fashion. The challenge lies in automatically identifying instances of the same folding pattern in different subjects, while not forcing correspondences (as they may not exist or vary significantly in some subjects). In particular, the goal is to automatically detect multiple distinct modes of appearance of particular cortical folds, in order to expose patterns of variability. This leads to several advantages over current approaches. Firstly, this permits the development of a bottom-up, parts-based description of cortical folds in healthy brains that copes with the inherent inter-subject variability. Secondly, uncovering this variability can also be helpful for numerous clinical applications by revealing which cortical folding patterns are common or rare across a subset of the population defined by pathology (e.g. schizophrenia, autism, some forms of epilepsy) versus the norm. The automatic, bottom-up nature of this system makes it possible to extract meaningful patterns, thus potentially leading to breakthroughs

in our understanding of neurological diseases, without requiring particular prior knowledge regarding the disease in question.

Many powerful local features have been developed in the field of computer vision, particularly for application areas such as object detection or classification. In this paper, we wish to explore the application of established features from the computer vision literature to this domain for the first time. In particular, we explore the power of different specifically chosen local image features to represent cortical folding patterns. Using machine learning techniques, a feature-based model is built to describe brain images as a collage of independent, co-occurring, distinct, localized image features which may not be present in all subjects. This approach is completely bottom-up, which does not restrict the discovery of variability to explicitly extracted sulcal structures, but rather automatically exposes patterns in the image independently of knowledge from particular anatomical structures. Three well-established feature types are explored (SIFT, KLT, Harris-affine), and their usefulness in learning cortical folding patterns is reported. By comparing them, the importance of selecting context-specific descriptions is emphasized. We also apply this model to hemisphere classification, in order to identify patterns of lateralization, and identify the usefulness of different feature types for this task.

The remainder of this paper is organized as follows. Section 2 describes the various local feature types used and reviews the modelling algorithm. Section 3 presents experimental results and discussions on modelling 196 lateral volume renderings of the International Consortium of Brain Mapping (ICBM) [11] data set for the tasks of identifying folding patterns and classifying hemispheres.

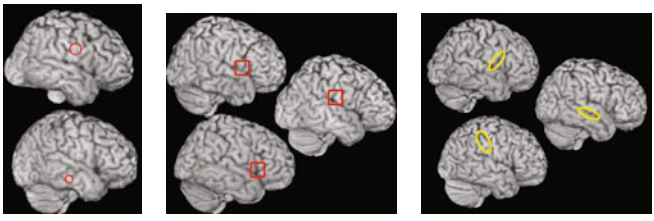
2 Feature-Based Modelling

2.1 Local Image Features

Local image features are distinct image patterns that can be detected automatically and robustly based on a saliency criterion. Depending on the saliency criterion of choice, an extracted local feature $f = (g, a)$ can be characterized geometrically by parameters including position, scale, orientation and eccentricity $g = (x, \sigma, \theta, e)$ of the local region. Its appearance is described by a 128-dimension vector a, storing gradients computed within the local region.

We use three different types of local features, and compare their performance in accurately describing cortical structures. We note that their relevance to the context plays a significant role in the usefulness of the model, both in describing brains as a collage, and in identifying class-distinctive features.

1. **SIFT:** Lowe's Scale invariant feature transform (SIFT) [12] evaluates saliency in the Gaussian scale-space. Although Gaussian scale-space maxima is effective for a wide range of applications, because it is generic enough to extract local features when there is no prior on the features of interest, it has had limited success with images of the cortex. SIFT generally extracts salient blob-like features, corresponding in this context to specular gyral reflections. See Fig 1(a) for examples.



(a) (b) (c)

Fig. 1. Examples of extracted features of each type, (a) SIFT, (b) KLT (c) Harris-affine

2. **KLT:** The Kanade-Lucas-Tomasi (KLT) [13] feature tracker defines saliency of local image patches using the Harris corner measure with uniform weighting function. Corner detectors are relevant for the cortex, as sulcal junctions and intersections are folding patterns of interest. See Fig 1(b) for examples.
3. **Harris-affine detector:** Mikolajczyk and Schmid's affine invariant keypoint detector [14] uses the Harris-Laplace transform for initial region point localization, and is robust to affine transformations by describing affine regions. The scale-invariant, elliptical affine regions are useful in describing elongated structures of interest, e.g. long gyri/sulci, of the cortex. See Fig 1(c) for examples.

2.2 Learning a Model

The subject images are assumed to be aligned after a pre-processing normalization step consisting of a 12-parameter affine image alignment, rather than a non-linear registration to a common reference frame. Local image features are extracted and stored for each subject image.

Learning a model effectively consists of identifying a set of model features $\{m_i\}$, consisting of cluster centroids of the very large set of extracted local image features $\{f_i\}$ [9] by identifying for each f_i, others which are similar in geometry and appearance. In fact, two different clusters G_i, A_i are constructed based on similarity in geometry and appearance respectively as follows:

$$G_i = \{f_j : \frac{\|x_i - x_j\|}{\sigma_i} < \varepsilon_x, \log|\frac{\sigma_j}{\sigma_i}| < \varepsilon_\sigma, \log|\frac{\theta_j}{\theta_i}| < \varepsilon_\theta, \frac{|e_i - e_j|}{e_i} < \varepsilon_e\}, \quad (1)$$

$$A_i = \{f_j : \|a_i - a_j\| < \varepsilon_{a_i}\} \text{ where } \varepsilon_{a_i} = \sup\{\varepsilon_a > 0 : \frac{A_i(\varepsilon_a) \cap G_i}{A_i(\varepsilon_a) \cap \bar{G}_i} > 1\}. \quad (2)$$

Given A_i, G_i, the model cluster associated with f_i is $M_i = A_i \cap G_i$, i.e. corresponding to similarity in both appearance and geometry. Features with strong appearance similarity are allowed to cluster together within a lenient geometrical window. Moreover, given similar geometry, features with significant appearance dissimilarity aren't forced to cluster together, thus allowing for multiple appearances modes at any particular location on the cortex.

Since several f_j (specifically those in M_i) will generate quasi-identical model clusters, the cluster set $\{M_i\}$ is pruned to remove this redundancy. Pruning consists of automatically removing, for all i, clusters M_j such that $|M_j| < |M_i|$ and $|M_i \cap M_j| > \varepsilon_M |M_j|$, where ε_M is an empirically determined pruning ratio. Additionally, degenerate clusters consisting of few features are also deleted. Finally, the set of model features is achieved by computing the centroids of model clusters $m_i = \langle M_i \rangle$.

The model features $\{m_i\}$ give insight to inherent group related variability of anatomical structures, as any particular feature is not necessarily found in all images. Rather, each subject's brain is modelled as an independent mosaic of local features, similar to multi-atlas approaches.

Furthermore, our feature-based modelling can automatically uncover this variability with a secondary clustering/pruning step on the set of model features. Features $m_{i_k}, k = 1, 2, ..., N$ with similar geometry (as defined in Eq. 1) are grouped together if they are also similar in appearance (as defined in Eq. 2, using a global ε_a), and averaged: $m_i = \langle \{m_{i_k}\}_k \rangle$. This removes multiple model feature representations of the same or similar image feature, and ideally reduces the number of model features at any particular location to a few distinct ones. In the context of sulci, these are representative of group-related variability of the cortex, corresponding to possible multiple distinct modes of appearance of particular cortical folds arising across a population.

3 Experimental Results and Discussion

3.1 Identifying Cortical Folds

Our first goal is to explore the cortical variability over a population of healthy subjects. For this reason, we chose to extract volumetric brain image data from the ICBM dataset [11]. It is common to first perform a continuous deformation, such as a spherical [16] or conformal [17] mapping, of the extracted cortical surface prior to looking at cerebral sulci. Rather than projecting the cortex onto a sphere, we chose to use a lateral projection. While this mapping is not a homeomorphism, it preserves the appearance of folding patterns with minimal perspective distortion within the region of interest.

Using the MRIcro software after applying the brain extraction tool, lateral renderings of the cortical surface, i.e. intensities in 2D lateral volume renderings defined by perspective projection, were outputted in image format. 98 unique subjects are used, mirroring hemispheres to obtain 196 images of resolution 217x181. In lateral views, the voxel resolution is 1mm isotropic, which is more than sufficient to visualize gyral/sulcal patterns in the laterally oriented surfaces.

Local features (KLT, SIFT, and Harris-affine) are automatically extracted from these images. KLT features are quite rare, averaging 50 per image, while SIFT and Harris-affine detectors can extract 400-500 features per image. Foreshortening along the edges of the brain due to projection results in lower resolution on the edges of the brain, therefore these features are disregarded. Additionally, all features are filtered by scale to remove extractions of small

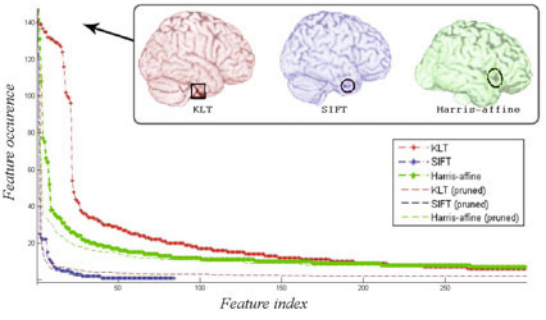


Fig. 2. Frequency of occurrence plot, illustrating the most frequent model features for each type

specular reflections and large structures, as the features of interest are at a characteristic scale determined by gyri/sulci dimensions. A feature-based model is built, as described in Section 2.2, separately for each feature type.

Figure 2 illustrates the frequency of model feature occurrences per type. The relatively small SIFT model size seems to imply that this generic feature does not have the matching capabilities of both KLT and Harris-affine features. Note that the most frequent features of each type correspond to various anatomical structures with low variability, tuned to the feature saliency criteria. As seen in Figure 2, the KLT corner detector reliably identifies the strong corner arising due to the cerebellum (shown in red), while the Harris-affine detector seems to extracts the slightly more elongated tri-corner of the pars triangularis in Broca's area (shown in green). In general, SIFT detects specular reflections off gyri, but its most frequent feature coincides with a fold near the mid temporal sulcus (shown in blue).

Figure 3 illustrates examples of model features for each type: occurrence matches are shown for two different, typical SIFT (S_1, S_2), four KLT (K_1, K_2, K_3, K_4), and four Harris-affine (H_1, H_2, H_3, H_4) model features across nine arbitrary subjects. The KLT tracker and Harris-affine detector are able to identify plausible cortical structures much better than the generic SIFT. KLT accurately detects strong corners arising from sulcal junctions, while the complementary Harris-affine features often coincide with elongated structures such as linear sulci and gyri edges. Although expert validation is needed to claim that these learned features correspond to valid semantic structures of the cortex, current results are encouraging for the context-specific features (KLT, Harris-affine), especially considering that matches are made despite significant anatomical variability. For example, H_1 is found at varying geometrical positions (see Figure 3), but seems to consistently correspond to the superior temporal gyrus. Similarly, H_4 accurately captures the length of the lateral sulcus and superior temporal sulcus, despite the variability in appearance, orientation and scale of these structures.

Additionally, the system automatically identifies examples of local regions of the cortex with at least two distinct modes of appearance, using the secondary

pruning step described in Section 2.2 based on a KLT model. Figure 4 shows three examples of such regions, with the corresponding appearance modes shown as image gradients. The distinct modes could correspond to the presence/absence of bridges or links between neighbouring gyri in parts of the population. It is noteworthy that, under inter-subject correspondence assumptions, these folding patterns would have been merged, and the underlying group-related variability would be lost in the registration.

3.2 Hemisphere Classification

In order to validate that the features we are examining do actually correspond to meaningful anatomical patterns, we now look at how they can be used to extract group-based differences. For the purposes of this paper, we explore the example toy application of hemisphere classification based on cortical folding patterns. The goal is not to establish an optimal strategy for hemisphere classification. Indeed, simpler methods can be successfully employed to address this particular task. Instead, the aim is to explore how well the particular features chosen can be used to automatically identify patterns of sulcal folds distinctive to a hemisphere, possibly indicative of lateralization of the brain.

Hemisphere classification is done for 98 subjects using 20-fold cross validation. Subject images (two hemispheres per subject) are either both in the training set or both in the test set, to avoid errors due to possible subject-specific feature correlation. Otherwise, training and test sets are randomly created. In Table 1, hemisphere classification error rates are shown for models based on three local feature types (KLT, SIFT, Harris-affine) using a support vector machine (SVM). Although our error rates are not comparable to other methods [15] which are specifically tailored to this problem, these results reveal the potential of classification based on various image features extracted from lateral renderings in a completely bottom-up fashion, without prior information about the problem context. In particular, tailoring features to the context seems to increase the performance of a classification model.

Table 1. 20-fold Cross Validation Error

-	**KLT**	**SIFT**	**Harris-affine**
SVM Error	0.24	0.45	0.33

While classification results are promising, our main goal is to explore how class-distinctive model features can be established automatically. Figure 5 shows the statistically significant class-distinctive model features, along with the corresponding contingency tables and significance values (note $p \ll 0.05$). The results of hemisphere classification indicate that it is likely not particular cortical folds that are most hemisphere-specific, but rather the overall geometry. The most distinguishing KLT model features, as shown in Fig 5(a), are corners that occur on and around the cerebellum, indicating that the size and shape of the temporal

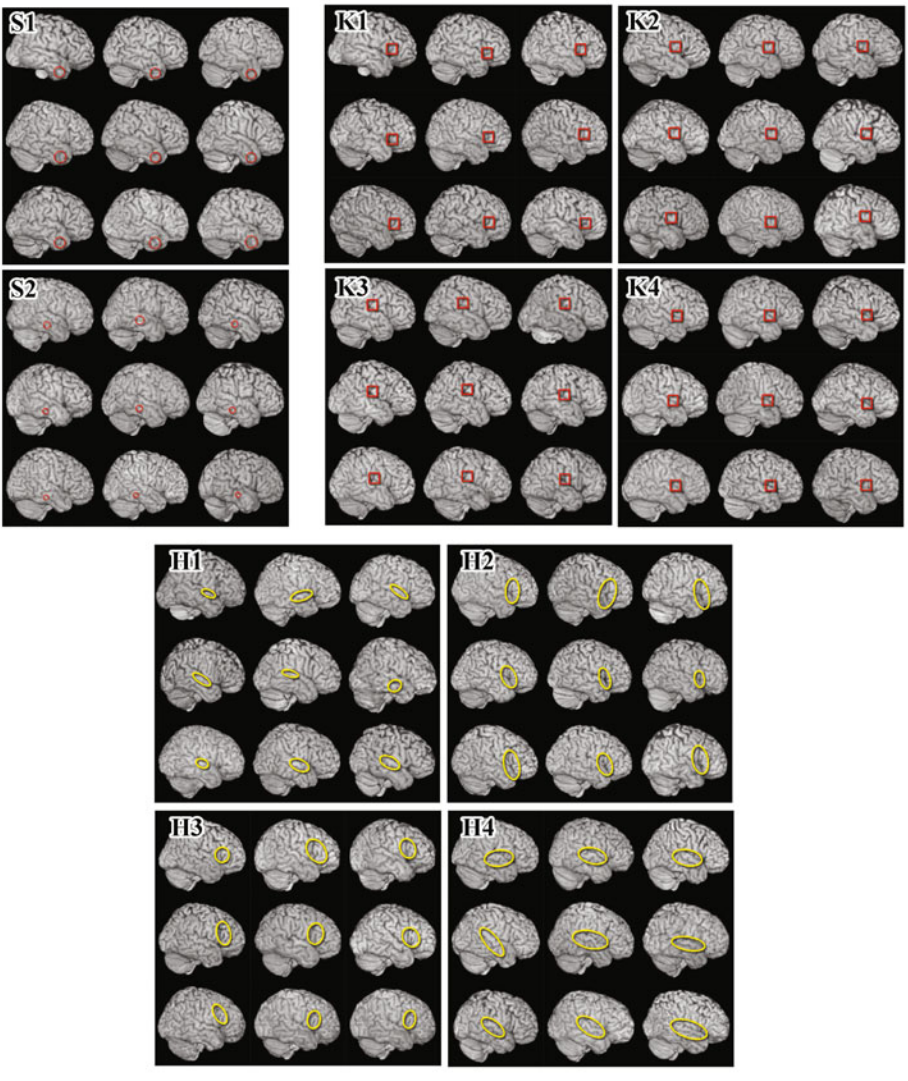


Fig. 3. Each grouping of nine images corresponds to occurrence matches of a particular model feature. Two different SIFT S_1, S_2 (red circles), four KLT K_1, K_2, K_3, K_4 (red squares), and four Harris-affine H_1, H_2, H_3, H_4 (yellow ellipses) features are shown.

lobe, which can obstruct this feature, is related to lateralization. This geometric warp between hemispheres is well studied [18]. However, sulcal folds can also be hemisphere-specific: the most distinguishing Harris-affine model features seem to occur on the pars triangularis of the inferior frontal gyrus (see Fig 5(b)), which is known to be larger on the left hemisphere [19].

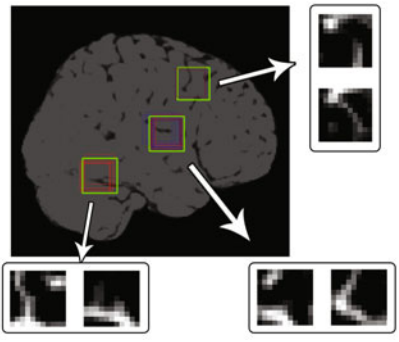
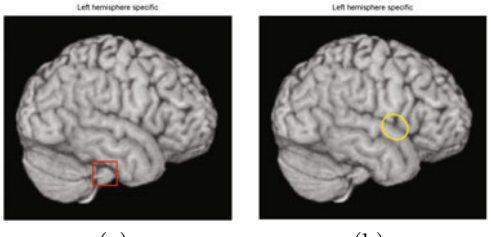


Fig. 4. Automatically found multiple distinct modes of appearances: Three local regions are shown on a brain template, with corresponding appearance modes shown as image gradients of folds



(a) (b)

	m_i	$\overline{m_i}$
Right	12	86
Left	37	61

$p = 2.7783 \times 10^{-5}$

(c)

	m_i	$\overline{m_i}$
Right	65	33
Left	89	9

$p = 2.1535 \times 10^{-5}$

(d)

Fig. 5. Significant class distinctive model features (a) found on the cerebellum (KLT) and (b) near the pars triangularis of Broca's area (Harris). Corresponding contingency tables and p-values are shown in (c), (d).

4 Conclusion

The contribution of this paper is a feature-based model of the cortex which aims to automatically uncover unlabelled cortical folding patterns from a large set of subject brain images. By describing brain images as a collection of independent, co-occurring, distinct, localized image features, the feature-based model is able to identify instances of the same folding pattern in different subjects while not forcing correspondences. Consequently, the model can automatically uncover group-related variability as well as multiple distinct modes of appearance of particular cortical folds. This will be helpful in complementing the laborious,

manual studies by neuroanatomists is developing an understanding of the the cortical variability of healthy brains, whose goal is to develop a general functional atlas of the brain. Uncovering this variability can be helpful for numerous clinical applications, by giving insight into which cortical folding patterns are common or rare across a subset of the population defined by pathology, for example. For instance, looking at brains of subjects with schizophrenia or autism may reveal differences in expression of particular folds with respect to the norm. This could be potentially instrumental in permitting us to learn about diseases where such variability is difficult to establish. As well, it could then be used as a means of evaluating similarity of a new subject to these pathological groups.

Various feature types (SIFT, KLT, Harris-affine) were explored in this work, and resulting models were also applied for hemisphere classification, exposing the power of different features and classifiers for this task. The class-distinctive features that were discovered may be indicative of brain lateralization. However, expert validation on the learned features is a necessary future step to confirm this. Interesting future work will involve applying this framework to subjects with schizophrenia or Alzheimer's, in order to analyze the class-distinctive features for this group. Future work could also involve the conjunction of features optimized for the task of cortical modelling and classification.

References

1. Ono, M., et al.: Atlas of the Cerebral Sulci. Thieme Medical (1990)
2. Toga, A.W., et al.: Probabilistic approaches for atlasing normal and disease-specific brain variability. Anat. Embryol., 267–282 (2001)
3. Mangin, J., et al.: Object-Based Morphometry of the Cerebral Cortex. IEEE TMI (2004)
4. Mangin, J. et al.: A framework to study the cortical folding patterns. NeuroImage (2004)
5. Klein, A. et al.: Mindboggle: a scatterbrained approach to automate brain labeling. NeuroImage (2005)
6. Blezek, D.J., Miller, J.V.: Atlas stratification. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 712–719. Springer, Heidelberg (2006)
7. Baloch, S., et al.: An anatomical equivalence class based joint transformation-residual descriptor for morphological analysis. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 594–606. Springer, Heidelberg (2007)
8. Toews, M., Arbel, T.: A Statistical Parts-based Appearance Model of Anatomical Variability. In: IEEE TMI, pp. 497–508 (2007)
9. Toews, M. et al.: Feature-Based Morphometry: Discovering Group-related Anatomical Patterns. NeuroImage (2009)
10. Toews, M., et al.: Automatically Learning Cortical Folding Patterns. In: IEEE ISBI, pp. 1330–1333 (2009)
11. Mazziotta, J., et al.: A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philos. Trans. R Soc. Lond. B Biol. Sci. 356(1412), 1293–1322 (2001)
12. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. IJCV 60(2), 91–110 (2004)

13. Tomasi, C., Shi, J.: Good Features to Track. In: CVPR, pp. 593–600 (1994)
14. Mikolajczyk, K., et al.: Scale and affine invariant interest point detectors. IJCV 60(1), 63–86 (2004)
15. Duchesnay, E., et al.: Classification from cortical folding patterns. IEEE TMI 26(4), 553–565 (2007)
16. Xu, C., et al.: A Spherical Map for Cortical Geometry. In: Proc. Int. Conf. Functional Mapping Human Brain, pp. 73–74 (1998)
17. Angenent, S., et al.: On the Laplace-Beltrami operator and brain surface flattening. IEEE Trans. Med. Imag. 18(8), 700–711 (1999)
18. Lyttelton, O., et al.: Positional and Surface Area Asymmetry of the Human Cerebral Cortex explored through automated surface-based analysis. Neuroimage (2009)
19. Foundas, A.L., et al.: Pars triangularis asymmetry and language dominance. Proc. Natl. Acad. Sci. USA 93(2), 719–722 (1996)

Surgical Phases Detection from Microscope Videos by Combining SVM and HMM

Florent Lalys[1,2,3], Laurent Riffaud[4], Xavier Morandi[1,2,3,4], and Pierre Jannin[1,2,3]

[1] INSERM, U746, Faculté de Médecine CS 34317, F-35043 Rennes Cedex, France
[2] INRIA, VisAGeS Unité/Projet, F-35042 Rennes, France
[3] University of Rennes I, CNRS, UMR 6074, IRISA, F-35042 Rennes, France
[4] Department of Neurosurgery, Pontchaillou University Hospital, F-35043 Rennes, France

Abstract. In order to better understand and describe surgical procedures by surgical process models, the field of workflow segmentation has recently emerged. It aims to recognize high-level surgical tasks in the Operating Room, with the help of sensors or human-based systems. Our approach focused on the automatic recognition of surgical phases by microscope images analysis. We used a hybrid method that combined Support Vector Machine and discrete Hidden Markov Model. We first performed features extraction and selection on surgical microscope frames to create an image database. SVMs were trained to extract surgical scene information, and then outputs were used as observations for training a discrete HMM. Our framework was tested on pituitary surgery, where six phases were identified by neurosurgeons. Cross-validation studies permitted to find a percentage of detected phases of 93% that will allow the use of the system in clinical applications such as post-operative videos indexation.

Keywords: Surgical phase, digital microscope, neurosurgery, SVM, HMM.

1 Introduction

In recent years, due to the progress of medicine and computers, there has been an increased use of technologies and in the Operating Room (OR). To develop computer assisted systems that better handle and integrate this new OR [1], a more detailed comprehension of the surgical workflow is needed. From this area, the challenge of surgical workflow recovery has emerged. Clinical applications are the evaluation of surgeons, OR management optimization or the creation of context-sensitive user interfaces. As mentioned in [2], the modelling must address behavioural, anatomical, pathological aspects and instruments. The concepts of Surgical Process (SP) and SP models (SPM) have been introduced for such purposes [2,3].

Related data extraction techniques can be classified according to the addressed level of granularity where the surgery is studied. These approaches yield to various methods used for data acquisition: patient specific procedures description [2,3], interview of surgeons [4], sensor-based methods [5-14], using fixed protocols created by expert surgeons [15], or combination between them [16]. Within sensor-based methods, an approach for finer classification is to differentiate materials used: such as a robot-simulator in virtual environments [5], using existing or additional materials.

B. Menze et al. (Eds.): MICCAI 2010 Workshop MCV, LNCS 6533, pp. 54–62, 2011.

Most of studies used sensors additionally installed. Padoy et al. [6] positioned sensors on instruments, and the workflow was segmented using Dynamic Time Warping (DTW) and Hidden Markov Model (HMM). Similarly, mechanisms for dataset preprocessing using Bayesian network before HMM training were presented in [7]. In both works, data acquisition was performed manually. Accelerometers placed on the operator were used in [8] to identify alphabets of activity. James et al. [9] installed an eye-gaze tracking system on surgeons combined with visual features to detect one important phase. Nara et al. [10] introduced an ultrasonic location aware system that tracks 3-D positions of the staff for the identification of surgical events.

From existing sensors within the OR, videos are a rich source of information, as demonstrated on laparoscopy [11]. A situation recognition process was created based on augmented reality and computer vision techniques. Helpful information such as 3D map were also extracted from laparoscopic videos in [12]. Bhatia et al. [13] analyzed global view videos for better OR management. Finally, Xiao et al. [14] implemented a system that record patient vital signs to situate the intervention process.

Our project focused on the extraction of information from microscope videos for high-level tasks recognition. Compare to other techniques, it permits not only to avoid the installation of materials, but also to have a source of information that has not to be controlled by human. Even if the approach and the application differ, we followed a methodology similar to the one described in [13] along with results of our previous work [17], where we presented in detail the image feature extraction process and performed studies on machine learning algorithms. Here the goal was to add a temporal reasoning for better detection. That's why we first took advantage of the ability of SVMs as binary classifiers to extract scene information from frames. Then outputs of the classification were treated as observations to train a HMM. This combination permitted to take into account the sequential nature of high-level tasks for accurate recognition. We focused in this paper on the detection of surgical phases and validated our methodology with a specific type of neurosurgical interventions: the pituitary surgeries.

2 Materials and Methods

The process for automatic recognition is introduced here: frames were first extracted from microscope videos, and reduced with spatio-temporal downsampling to perform feature extraction. Image signatures were composed of 185 features, in which discriminant ones were chosen with a specific feature selection. SVMs were then used to classify relevant surgical scene information. These results alone were not enough informative to correctly classify surgical phases, so a HMM has then been trained, taking as observations the outputs of the SVMs and as hidden states the surgical phases. The Viterbi decoder finally permitted to find the optimal path for a given observation sequence. We assessed this process with cross-validation studies.

2.1 Data-Set

We evaluated our algorithm on pituitary surgeries [18], presented in [17]. Pituitary adenomas are tumors that occur in the pituitary gland, where neurosurgeons use a

trans-nasal approach with an incision in the back wall of the nose. The anonymous data set was composed of 16 pituitary surgeries (mean time: 40min), all performed in the neurosurgical department of Rennes University Hospital by three expert surgeons. Videos focused on the operative field of view and were recorded using the surgical microscope OPMI Pentero (Carl Zeiss) (Videos: 768 x 576 pixels at 33 frames per second). The labeling of surgical phases was manually performed by surgeons (Fig. 1.).

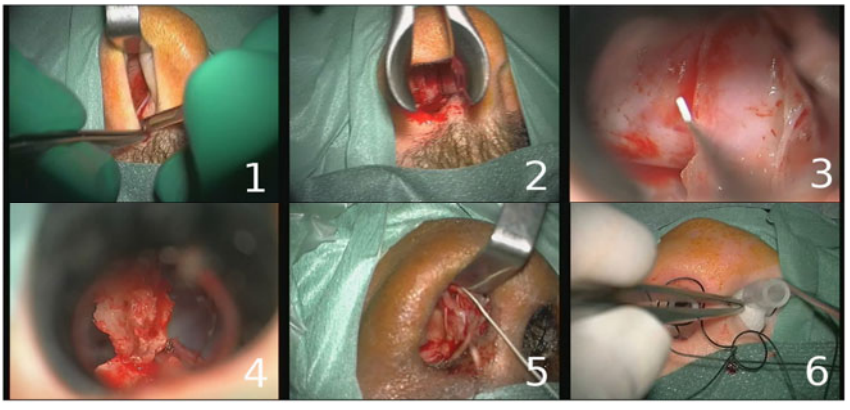


Fig. 1. Example of typical digital microscope images for the six phases: 1) nasal incision, 2) nose retractors installation, 3) access to the tumor along with tumor removal, 4) column of nose replacement, 5) suturing, 6) nose compress installation

Original frames were first spatially downsampled by a factor of 8 with a 5-by-5 Gaussian kernel (internal studies have shown that it had no impact on accuracy) and then downsampled to 1 frame every 2s (0.5Hz). We performed a statistical intensity normalization of images, where data closely followed a normal distribution.

2.2 Feature Extraction and Selection

From these videos, we randomly extracted 500 frames which were supposed to correctly represent the six phases of a common pituitary surgery. We defined for each frame a feature vector, representing a signature. Signatures are composed of three main information that usually describe an image: the color, the texture and the form.

The color has been extracted with two complementary spaces [19]: RGB space (3 x 16 bins) along with Hue (32 bins) and Saturation (32 bins) from HSV space. The texture has been extracted with the co-occurrence matrix along with Haralick descriptors [20]. The form was represented with spatial moments [21], and we also computed the Discrete Cosine Transform (DCT) [22] coefficients. Each signature was finally composed of 185 complementary features.

The main goal of feature selection is to remove redundancy information and to keep essential ones. Feature selection methods can be divided into two groups [23]: the filter and the wrapper approach. We fused them using the method described by Mak and Kung [24], where they argued that both methods are complementary to each other. Algorithms were first independently applied to find two feature subsets. They

were then merged by selecting one feature at a time from the two subsets, starting from the highest rank. The Recursive Feature Elimination (RFE) SVM [25] was chosen for wrapper method. The principle is to generate the ranking using backward feature elimination. The mutual information (MI) [26] was chosen for the filter method, where a feature is more important if the MI between the target and the feature distributions is larger. In order to a have a good compromise between computation time and accuracy, we kept the 40 first features.

2.3 Supervised Classification

Based on our last work [17], multiclass SVMs [27] have been found to be effective for microscope images classification. We decided to use binary SVMs for scene information extraction. SVMs are supervised learning algorithms used for classification and regression. Mathematically, given training data $\{x_1...(x_n)\}$where $x \in \Re^d$ and their labels $\{y_1...(y_n)\}$where $y \in \{-1,(1)\}$. The goal is to find the optimal hyperplane $w.x+b=0$ that separates the data into two categories. The idea is to maximize the margin between the positive and negative examples. The parameter pair $(w;b)$ is finally the solution to the optimization problem:

$$\varphi(w)=\frac{1}{2}|w|^2=\frac{1}{2}(w.w) \tag{1}$$

following constraints:

$$y_i(x_i.w+b)-1\geq 0,\ \forall i \tag{2}$$

Four discriminant scene information were defined: global-zoom views, presence-absence of nose retractors, of the column of nose and of compress. Combinations of these 4 binary outputs resulted in 16 possible observations for the HMM.

2.4 HMM

HMMs [28] are statistical models used for modeling of non-stationary vector times-series. An HMM is formally defined by a five-tuple (S,O,Π,A,B), where $S=\{s_1...(s_N)\}$ is a finite set of N states, $O=\{o_1...(o_M)\}$ is a set of M symbols in a vocabulary, $\Pi=\{\pi(i)\}$ are the initial state probabilities, $A=\{a(ij)\}$ the state transition probabilities and $B=\{b_i(o(k))\}$the output probabilities. Here, outputs of SVMs were treated as observations for the HMM. States were represented by the surgical phases that generated a left-right HMM (Fig. 2.). The transition probabilities were low because of the sampling rate of frames (0.5Hz). We set two probabilities for the transition from one state to its consecutive state: $\alpha=0.005$ for state n°3, $\beta=0.01$for the others. The probability of remaining in the same state is then: $1-\alpha$ or $1-\beta$. The outputs probabilities were obtained from SVMs results. They were computed as the probability of having an observation in a specific state. Videos were applied to SVMs and observation probabilities were manually computed. All these probability

computations were part of the training process of our framework, and were therefore obtained only from the training sample of the cross-validation. Furthermore, the Baum-Welch algorithm has not been used for the training because of the limited size of the training sample.

Lastly, given observations and the HMM structure, the Viterbi algorithm [29] find the most likely sequence of states.

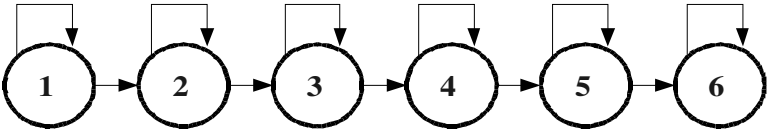


Fig. 2. Left-right HMM, where each state corresponds to one surgical phase

2.5 Cross-Validation

SVM classifiers and HMM were both evaluated with a random 10-fold cross-validation study. The image database and videos were divided into 10 random subsets. Nine were used for training while the prediction was made on the 10^{th} subset. This procedure was repeated 10 times and results were averaged. We computed the correct classification rate for SVMs evaluation and the percentage of phases misclassified, namely the Frequency Error Rate (FER), for HMM assessment. In addition, the confusion matrix was extracted, showing exactly where states were misclassified.

3 Results

Statistical results of the cross-validation study for the SVMs (Tab. 1.) showed that very good detections (~88%) along with low standard deviations (max=2.5%) were obtained, getting a maximum accuracy of 94.6% for the column of nose detection.

Table 1. Mean accuracy and standard deviation (Std) of the 4 scene information recognition

	Global-zoom view	Presence-absence of nose retractors	Presence-absence of column of nose	Presence-absence of compress
Accuracy (%)	87.6	88.0	94.6	87.4
Std (%)	2.4	2.2	1.6	2.5

The HMM study showed a **mean FER of 7.1 +/- 5.3%**, with a **min of 2.5%** and a **max of 15.2%.** This error rate is low, but values are very scattered (resulting in a high standard deviation). A recognized sequence compare to the true one is shown on Fig. 3. On this particular example, each state is correctly classified with a maximum delay of 40s.

From Tab. 2., we see that state n°3 contains the bigger number of frames, and all confusions are always between neighbouring states. The most significant error is for state n°5, where the detection is around 75%. The highest accuracy (excluding the first and the last state) is for state n°4, where the detection reaches 95%.

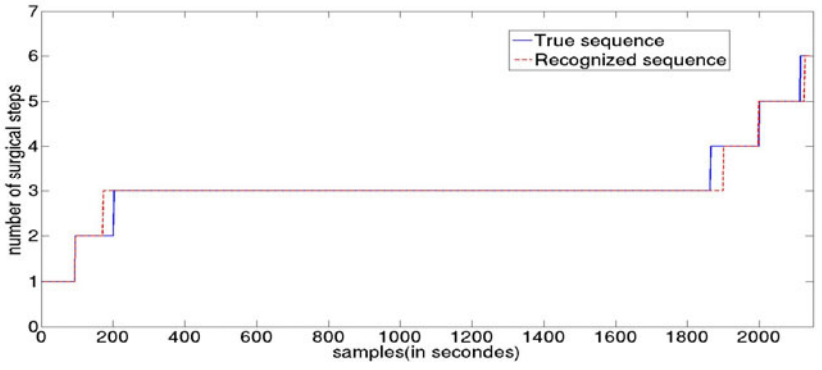


Fig. 3. Phases recognition of one video made by the HMM compared to the ground truth

Table 2. Confusion matrix for the surgical phases detection with the Viterbi decoder. Rows indicate the recognized surgical steps and columns the ground truth.

	1	2	3	4	5	6
1	5.68	0.97	0	0	0	0
2	0	4.68	4.09	0	0	0
3	0	0	72.99	0.12	0	0
4	0	0	0.45	3.12	0.07	0
5	0	0	0	0.04	3.31	0
6	0	0	0	0	0.99	3.49

4 Discussion

In this paper, we proposed a framework that automatically recognizes surgical phases from microscope videos. The combination of SVMs and HMM showed a total accuracy of 93% of detected phases.

4.1 Microscope Video Data

As mentioned in [13], the information extracted from the OR must be discriminant, invariant to task distortion, compact in size and easy to monitor. Microscope video data meet all of these constraints. Image features are first very discriminant for scene information extraction, as the SVMs validation indicated. Secondly, within a same surgical environment, procedures are reproducible and image features are thus invariant to task distortion. This constraint addresses the issue of the adaptability of the system. Due to the different equipments in each department, the system could be not flexible. The solution would be to train dedicated image databases for each department which would be adapted to the corresponding surgical environment and microscope scene layout. The idea would also be to have several models for each type of procedure, adapting the scene information to be extracted and optimising the HMM. The third crucial parameter is the sample size which must be compact. Image signatures are composed of 40 features and are thus reduced. Finally, the real added value is the use of microscope videos. This device is not only already installed in the OR, but it has also not to be monitored by the staff.

4.2 Accuracy of the Detection

We decided to use frames in a static way (without motion), and deliberately remained at a high level of granularity with the detection of phases. The recognition of lower level information, such as gestures, is difficult with microscope videos only. Spatio-temporal features will have to be inserted for the segmentation of such information.

The possible phases “access to the tumor” and “tumor removal” were fused because the transition between both was not clear due to similar tools and zooms. The confusion matrix showed that there was no main confusion and that the HMM was helpful for separating phases with same image features (like phase n°1 and n°5).

The high recognition rates of binary SVMs, associated with their small standard deviations, indicates that they are very robust for images classification. Then graphical probabilistic models allow an efficient representation of the problem by modelling time varying data. This association permitted to obtain good detection accuracy.

4.3 Clinical Applications

Workflow recovery might be helpful for various applications. Purposes are generally to bring a plus-value to the surgery or to the OR management. This work could be integrated in an architecture that would extract microscope frames and transform it into information helping the decision making process. For now, even with a low computation time (feature extraction/selection + classification < 0.5s), accuracy must definitively be higher than our results before establishing on-line surgical phase detection.

However, the system could be introduced for video indexation. Surgical videos are very useful for learning and teaching purposes, but surgeons often don't use them because of the huge amount of surgical videos. The data-base would contain relevant surgical phases of each procedure for easy browsing. Moreover, we could imagine the creation of pre-filled reports that will have to be completed by surgeons. For such applications, even with few errors, the automatic indexation would be helpful, as there is no need of perfect detection and it has no impact on the surgery.

5 Conclusion

Using the proposed framework, we are now able to recognize the major surgical phases of every new procedure, by computing frames signatures, classifying scene information, and decoding SVMs outputs with the Viterbi algorithm. Thanks to this combination, we obtained high detection accuracy. We have validated this framework with pituitary surgeries. Six phases were defined by an expert, and we found a global accuracy of 93% of detected phases. This recognition process is a first step toward the construction of context-aware surgical systems. Currently, it could be used for post-operative video indexation or reports generation. In future works, spatial image features will have to be mixed with other information (such as spatio-temporal features) to generate a more robust system.

Acknowledgments. The authors would like to acknowledge the financial support of Carl Zeiss Surgical GmbH.

References

1. Cleary, K., Chung, H.Y., Mun, S.K.: OR 2020: The operating room of the future. Laparoendoscopic and Advanced Surgical Techniques 15(5), 495–500 (2005)
2. Jannin, P., Morandi, X.: Surgical models for computer-assisted neurosurgery. Neuroimage 37(3), 783–791 (2007)
3. Neumuth, T., Jannin, P., Strauss, G., Meixensberger, J., Burgert, O.: Validation of Knowledge Acquisition for Surgical Process Models. J. Am. Med. Inform. Assoc. 16(1), 72–82 (2008)
4. Morineau, T., Morandi, X., Le Moëllic, N., Diabira, S., Haegelen, C., Hénaux, P.L., Jannin, P.: Decision making during preoperative surgical planning. Human Factors 51(1), 66–77 (2009)
5. Darzi, A., Mackay, S.: Skills assessment of surgeons. Surg. 131(2), 121–124 (2002)
6. Padoy, N., Blum, T., Feuner, H., Berger, M.O., Navab, N.: On-line recognition of surgical activity for monitoring in the operating room. In: Proc. of IAAI (2008)
7. Bouarfa, L., Jonker, P.P., Dankelman, J.: Discovery of high-level tasks in the operating room. Journal of Biomedical Informatics (in Press, 2010)
8. Ahmadi, S.A., Padoy, N., Rybachuk, K., Feussner, H., Heinin, S.M., Navab, N.: Motif discovery in OR sensor data with application to surgical workflow analysis and activity detection. In: M2CAI Workshop, MICCAI, London (2009)
9. James, A., Vieira, D., Lo, B.P.L., Darzi, A., Yang, G.-Z.: Eye-gaze driven surgical workflow segmentation. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 110–117. Springer, Heidelberg (2007)
10. Nara, A., Izumi, K., Iseki, H., Suzuki, T., Nambu, K., Sakurai, Y.: Surgical workflow analysis based on staff's trajectory patterns. In: M2CAI Workshop, MICCAI, London (2009)
11. Speidel, S., Sudra, G., Senemaud, J., Drentschew, M., Müller-stich, B.P., Gun, C., Dillmann, R.: Situation modeling and situation recognition for a context-aware augmented reality system. Progression in Biomedical Optics and Imaging 9(1), 35 (2008)
12. Sánchez-González, P., Gayá, F., Cano, A.M., Gómez, E.J.: Segmentation and 3D reconstruction approaches for the design of laparoscopic augmented reality environments. In: Bello, F., Edwards, E. (eds.) ISBMS 2008. LNCS, vol. 5104, pp. 127–134. Springer, Heidelberg (2008)
13. Bhatia, B., Oates, T., Xiao, Y., Hu, P.: Real-time identification of operating room state from video. In: AAAI, pp. 1761–1766 (2007)
14. Xiao, Y., Hu, P., Hu, H., Ho, D., Dexter, F., Mackenzie, C.F., Seagull, F.J.: An algorithm for processing vital sign monitoring data to remotely identify operating room occupancy in real-time. Anesth Analg. 101(3), 823–832 (2005)
15. MacKenzie, C.L., Ibbotson, A.J., Cao, C.G.L., Lomax, A.: Hierarchical decomposition of laparoscopic surgery: a human factors approach to investigating the operating room environment. Min. Invas. Ther. All Technol. 10(3), 121–128 (2001)
16. Neumuth, T., Czygan, M., Goldstein, D., Strauss, G., Meixensberger, J., Burgert, O. Computer assisted acquisition of surgical process models with a sensors-driven ontology. In: M2CAI Workshop, MICCAI, London (2009)
17. Lalys, F., Riffaud, L., Morandi, X., Jannin, P.: Automatic phases recognition in pituitary surgeries by microscope images classification. In: Navab, N., Jannin, P. (eds.) IPCAI 2010. LNCS, vol. 6135, pp. 34–44. Springer, Heidelberg (2010)

18. Ezzat, S., Asa, S.L., Couldwell, W.T., Barr, C.E., Dodge, W.E., Vance, M.L., McCutcheon, I.E.: The prevalence of pituitary adenomas: a systematic review. Cancer 101(3), 613–622 (2004)
19. Smeulders, A., Worrin, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Mach. Intell. 22(12), 1349–1380 (2000)
20. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. IEEE Trans. on Systems, Man, and Cybernetics 3(6), 610–621 (1973)
21. Hu, M.: Visual pattern recognition by moment invariants. Trans. Inf. Theory 8(2), 79–87 (1962)
22. Ahmed, N., Natarajan, T., Rao, R.: Discrete Cosine Transform. IEEE Trans. Comp., 90–93 (1974)
23. Duda, R.O., Hart, P.E.: Pattern classification and scene analysis. Wiley, New York (1973)
24. Mak, M.W., Kung, S.Y.: Fusion of feature selection methods for pairwise scoring SVM. Neurocomputing 71, 3104–3113 (2008)
25. Guyon, I., Weston, J., Barhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machine. Machine Learning 46, 389–422 (2002)
26. Hamming, R.W.: Coding and Information Theory. Prentice-Hall Inc., Englewood Cliffs (1980)
27. Crammer, K., Singer, Y.: On the Algorithm implementation of multiclass SVMs. JMLR (2001)
28. Rabiner, L.R.: A tutorial on Hidden Markov Models and selected applications in speech recognition. Proc of IEEE 77(2) (1989)
29. Viterbi, A.: Errors bounds for convolutional codes. IEEE TIT 13(2), 260–269 (1967)

Motion Artifact Reduction in 4D Helical CT: Graph-Based Structure Alignment

Dongfeng Han[2], John Bayouth[2], Sudershan Bhatia[2], Milan Sonka[1,2], and Xiaodong Wu[1,2]

[1] Department of Electrical and Computer Engineering
[2] Department of Radiation Oncology
The University of Iowa, Iowa City, IA, USA
{dongfeng-han,john-bayouth,sudershan-bhatia,
milan-sonka,xiaodong-wu}@uiowa.edu

Abstract. Four dimensional CT (4D CT) provides a way to reduce positional uncertainties caused by respiratory motion. Due to the inconsistencies of patient's breathing, images from different respiratory periods may be misaligned, thus the acquired 3D data may not accurately represent the anatomy. In this paper, we propose a method based on graph algorithms to reduce the magnitude of artifacts present in helical 4D CT images. The method strives to reduce the magnitude of artifacts directly from the reconstructed images. The experiments on simulated data showed that the proposed method reduced the landmarks distance errors from 2.7 mm to 1.5 mm, outperforming the registration methods by about 42%. For clinical 4D CT image data, the image quality was evaluated by the three medical experts and both of who identified much fewer artifacts from the resulting images by our method than from those by the commercial 4D CT software.

1 Introduction

Four-dimensional (3D + time) multi-detector computed tomography (CT) imaging technology provides human body images at different respiratory phases while applied to the breathing lung [1,2]. This imaging approach, enabling the direct incorporation of organ motion into treatment planning, is extremely valuable for thoracic radiotherapy. Due to current spatio-temporal limitation of CT scanners, the entire body can not be imaged in a single respiratory period. One widely used method in clinic to acquire a 4D CT image of a patient is to use the CT scanner in helical mode, that is, image data for adjacent couch positions are continuously acquired in sequence. To obtain time-resolved image data during the periodic motion, multiple image slices must be reconstructed at each couch position for a time interval equal to the duration of a full respiratory period [2], which can be achieved in helical mode with very low pitch (i.e., the couch moves at a speed low enough so that a sufficient number of slices can be acquired for a full respiratory period). Because of the use of the multi-detector scanner, the 2D image slices acquired at each couch position form an image stack, which is associated with a measured respiratory phase and covers only part of the patient's

B. Menze et al. (Eds.): MICCAI 2010 Workshop MCV, LNCS 6533, pp. 63–73, 2011.

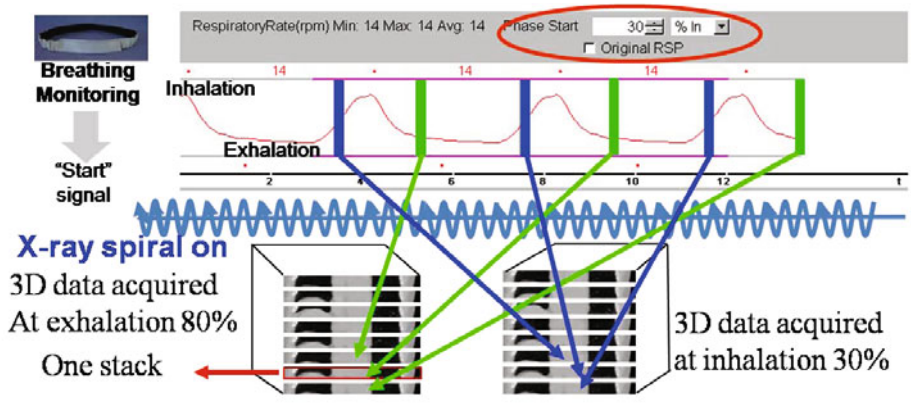
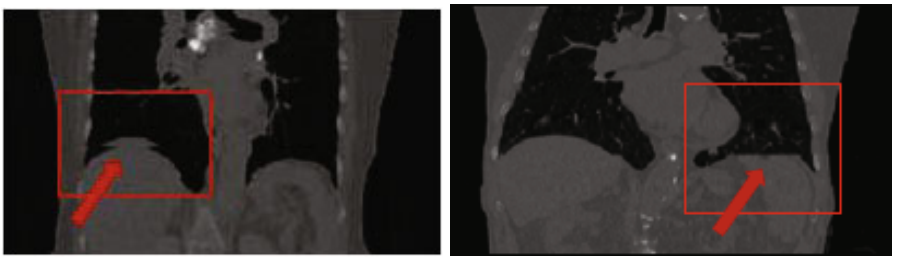


Fig. 1. An illustration of 4D CT imaging. 4D CT image data consists of a series of multiple 3D CT volume datasets acquired at different respiratory phases. Each phase-specific 3D CT dataset is made of several groups of 2D images (stacks), where each stack is reconstructed from each period of respiration during acquisition.

body. In the post-processing stage, the stacks from all the respiratory periods associated with a same specific measured respiratory phase are stacked together to form a 3D CT image of the patient for that phase. A 4D CT image is then reconstructed by temporally viewing the 3D phase-specific datasets in sequence (Fig. 1). It is also possible to acquire a 4D CT image using a multi-detector CT scanner in cine mode [1] (the couch stops during data acquisition).

However, due to the variability of respiratory motion, image stacks from different respiratory periods could be misaligned, causing the resulting 4D CT data does not accurately represent the anatomy in motion [3]. The artifacts can be categorized into two major types: anatomy overlap and anatomy gap [4]. As shown in Fig. 2(a), if the anatomies represented by two image stacks at the same measured respiratory phase from two consecutive periods partially overlap each other, the artifact of anatomy overlap appears while simply stacking them together during the 4D CT reconstruction. On the contrary, if the anatomies between those represented by the two stacks are missing, the stacking operation causes the artifact of anatomy gap (Fig. 2(b)).

All current 4D CT acquisition and reconstruction methods frequently lead to spatial artifacts; a recent study shows these artifacts occur with an alarmingly high frequency and spatial magnitude [3]. Therefore, significant improvement for reducing the artifacts is needed in 4D CT imaging. One way to solve the problem is to use deformable registration. However, general deformable registration methods are not structure-aware, causing structure inconsistency and further producing visual artifacts. The optimal seam detection methods was reported in Refs.[5,6] fail when the initial structures are not well aligned. In Refs.[7,8], the authors proposed methods to obtain 4D CT images with reduced artifacts for cine mode. In the cine mode acquisition, the raw projection data is acquired



(a) Anatomy overlap. (b) Anatomy gap.

Fig. 2. Two types of artifacts due to irregular respiration

over multiple periods of respiration for a given couch position, while in the helical mode the couch position for each period of respiration is unique. Thus, the method specific to the cine mode may not work for the helical mode data acquisition.

In this paper, we focus on the basic step in 4D CT image reconstruction, that is, how to stitch two image stacks S_i and S_j that partially overlap in anatomy, to obtain a spatio-temporally coherent data set, further reducing the artifacts. To our best knowledge, no method has definitively solved the anatomy gapping problem. By acquiring the raw projection data using a continuous x-ray source and couch motion, all the patient's anatomy is imaged and presented in at least one phase. Even when an anatomy gap occurs due to the irregular respiration for moving and/or deforming tissues, rigid and stationary anatomy presents within the image. Our method exploits the fact that anatomy gap is essentially avoidable. Several ways can help to achieve that: (1) we control either the pitch (couch translation speed and/or x-ray tube rotation speed) or the patient's respiratory rate (increasing breaths per minute by training) to guarantee the overlaps between stacks from adjacent couch positions to minimize anatomy gap; and (2) we utilize a bridge stack S_b acquired from another respiratory phase which overlaps with both image stacks. Then we perform the following stitching operations: first stitch S_i and S_b to obtain an intermediate stack S_{ib}, which is overlap with S_j; and then stitch S_{ib} with S_j resulting an artifact-reduced stack S_{ibj}. Thus, the fundamental problem is to stitch two partially overlapping stacks. However, in this case we need to be aware of the deformation errors induced by using the bridge stack from a different phase.

We propose a novel method based on graph algorithms for solving the stitching problem, in which the misalignment of the anatomy structures is substantially reduced. In order to achieve our goal, we first compute an interface seam for the stitching in the overlapping region of the first image stack, which passes through the "smoothest" region to reduce the structure complexity along the stitching interface. Then, the corresponding interface seam in the second image stack is computed using our proposed seam flow method, which essentially solves

a multiple-label problem in Markov Random Fields. The two image stacks are stitched along the interface seams based on the computed flow vector field.

2 Methods

Given two image stacks, I and I', with a partial overlap in anatomy, we want to stitch them together to form a spatially coherent image, i.e., to minimize the artifacts in the resulting image as much as possible. Assume that $\Omega \subset I$ overlaps with $\Omega' \subset I'$. We call I (resp., I') the fixed (resp., moving) image. Our method consists of the following five steps (Fig. 3): (1) Initial registration to roughly align I and I'; (2) Computing an interface seam in the fixed image to reduce the alignment ambiguity; (3) Finding the interface seam in the moving image by the seam flow method; (4) Propagation the flow to the rest of the moving image; (5) Warping the images and getting an artifact-reduced image. In this paper, we focus on the main steps (2), (3) and (4).

2.1 Computing the Optimal Interface Seam in Ω

To reduce the alignment ambiguity, the interface seam for stitching is required to pass through the "smoothest" region in Ω with less structure complexity. We model this problem as an optimal seam detection problem by the graph searching method [9], which has been widely used in different applications [10,11,12].

The Interface Seam. Recall that $\Omega(\mathbf{x}, \mathbf{y}, \mathbf{z})$ denotes the overlapping region of the fixed image I. We can assume that the size of Ω is $X \times Y \times Z$, and the two image stacks, I and I', are stitched along the $\mathbf{z}$-dimension. Thus, the interface seam S in Ω is orthogonal to the $\mathbf{z}$-dimension and can be viewed as a function $S(x, y)$ mapping (x, y) pairs to their z-values. To ease the stitching, we certainly hope the interface seam itself is smooth enough. We thus specify the maximum allowed changes in the $\mathbf{z}$-dimension of a feasible seam along each unit distance change in the $\mathbf{x}$- and $\mathbf{y}$-dimensions. More precisely, if $\Omega(x, y, z')$ and $\Omega(x, y + 1, z'')$ (resp., $\Omega(x, y, z')$ and $\Omega(x + 1, y, z'')$) are two neighboring

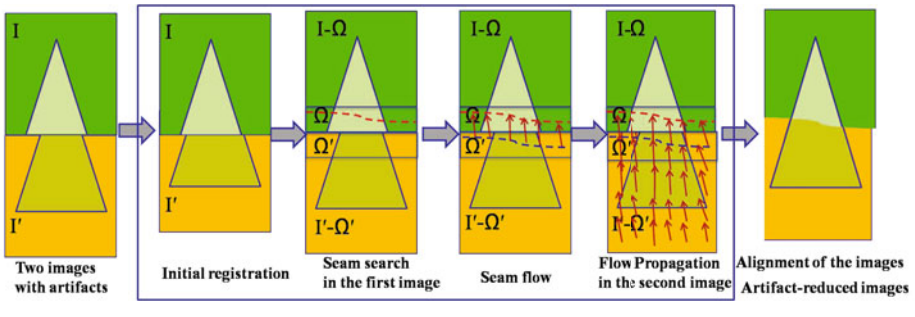


Fig. 3. Illustrating steps of the proposed method

voxels on a feasible seam and $\delta_{\mathbf{y}}$ and $\delta_{\mathbf{x}}$ are two given *smoothness parameters*, then $|z' - z''| \leq \delta_{\mathbf{y}}$ (resp., $|z' - z''| \leq \delta_{\mathbf{x}}$).

The Energy Function enforces the interface seam passing through the region of Ω with less structure complexity. Two factors should be considered: (1) The gradient smoothness in Ω which prevents the seam from breaking anatomy edges; and (2) the similarity between the neighboring voxels in the overlapping region of the fixed image (Ω) and that of the moving image (Ω'). Let $C_s(p)$ denote the gradient smoothness cost of the voxel at $p(x, y, z)$ and $C_d(p)$ be the dissimilarity penalty cost of voxel p under the neighborhood setting $\mathcal{N}$. Denote S a feasible interface seam. The energy function that needs to be minimized is defined by the following equation:

$$\mathcal{F}(S) = \alpha \sum_{p \in S} C_s(p) + (1 - \alpha) \sum_{p \in S} C_d(p), \tag{1}$$

where α is used to balance $C_s(p)$ and $C_d(p)$ and

$$C_s(p) = \sqrt{\left\| \frac{\partial \Omega(x, y, z)}{\partial x} \right\|^2 + \left\| \frac{\partial \Omega(x, y, z)}{\partial y} \right\|^2 + \left\| \frac{\partial \Omega(x, y, z)}{\partial z} \right\|^2}, \tag{2}$$

and

$$C_d(p) = (\Omega(p) - \Omega'(p))^2 + \sum_{(p,q) \in \mathcal{N}} (\Omega(q) - \Omega'(q))^2. \tag{3}$$

Optimization. The problem of finding in Ω an optimal interface seam S while minimizing the objective function $\mathcal{F}(S)$ is in fact an optimal single surface detection problem, which can be solved by computing a minimum-cost closed set in the constructed graph from Ω [9].

2.2 Seam Flow via Graph Cuts

The optimal interface seam in the moving image (actually, in the overlapping region Ω' of the moving image) is computed by the seam flow, which is achieved by solving a multiple-label problem in Markov Random Fields (MRFs).

Seam Flow as Graph Labeling. Intuitively, seam flow means to "move" the interface seam in the fixed image to the moving image to find the corresponding one, which is the "best" match with the one in the fixed image. We model it as a multiple labeling problem. A label assignment l_p to a voxel $p(x, y, z)$ on the seam S is associated with a displacement vector $\mathbf{f}_p = (f_x, f_y, f_z)$, called the *seam flow* of S. That is, we map $p(x, y, z) \in \Omega$ in the fixed image to $p'(x+f_x, f+f_y, z+f_z) \in \Omega'$ in the moving image. Thus, the problem is modeled as a multiple labeling problem, where each node corresponds to one voxel on the seam and the label for each node at position p is denoted by l_p. The energy $E(\mathcal{L})$ of a labeling $\mathcal{L}$ is the log-likelihood of posterior distribution of an MRF [13]. $E(\mathcal{L})$ is composed of a data term E_d and a spatial smoothness term E_s,

$$E(\mathcal{L}) = \sum_{p \in S} E_d(l_p) + \beta \sum_{p \in S, q \in S, (p,q) \in \mathcal{N}} E_s(l_p, l_q), \tag{4}$$

where $\mathcal{N}$ denotes the neighboring system and β is the parameter to balance the two terms. Suppose label l_p is defined by the displacement $\mathbf{f}_p$. The data term for each node on the interface seam S is defined, as follows.

$$E_d(l_p) = \frac{1}{(2w+1)^3} \sum_{i=-w}^{w} \sum_{j=-w}^{w} \sum_{k=-w}^{w} (I(x+i, y+j, z+k) - I'(x+i+f_x, y+j+f_y, z+k+f_z))^2, \tag{5}$$

where w is the window size. $E_d(l_p)$ is in fact the block matching score between $p \in \Omega$ and its corresponding voxel in Ω'. Assume that nodes $p(x, y, z)$ and $q(x', y', z')$ are adjacent on seam S under the neighborhood setting $\mathcal{N}$. The spatial smoothness of their labels is defined as,

$$E_s(l_p, l_q) = \sqrt{(f_x - f_{x'})^2 + (f_y - f_{y'})^2 + (f_z - f_{z'})^2}. \tag{6}$$

The spatial smoothness term helps preserve the structure of the seam S.

Approximation by Graph Cuts. The function defined in Eq. (4) leads to an energy minimization problem in MRF, which is computationally intractable (NP-hard). However, it has been shown in [14] that an approximate solution can be found that typically produces good results using the multiple label graph cuts method[13]. A hierarchical approach with three levels pyramid is used to reduce the computation complexity.

2.3 Flow Propagation by Solving Laplace Equation

Once an optimal seam flow $\mathbf{f}$ is computed on S, the flow needs to propagate to the rest of the moving image. To smoothly propagate the flow $\mathbf{f}$ to the whole moving image I', we minimize the following Laplace equation with Dirichlet boundary condition [6,15] to obtain the propagation flow $\mathbf{f}^*$.

$$\arg\min_{\mathbf{f}^*} \int\int\int_{I'} |\nabla \mathbf{f}^*|^2 \quad s.t. \mathbf{f}|_S = \mathbf{f}^*|_S, \tag{7}$$

where ∇ is a gradient operator. The problem can be discretized and solved by the conjugate gradients method.

Using the propagation flow $\mathbf{f}^*$ in I', we perform a warping with a bilinear interpolation in I', resulting an artifact-reduced image [6,15].

3 Experimentation and Results

3.1 Experimental Method and Material

Parameters Setting. $\delta_{\mathbf{x}}$ and $\delta_{\mathbf{y}}$ are the smoothness parameters. In our experiments we used $\delta_{\mathbf{x}} = \delta_{\mathbf{y}}$=4. In Eq.(1), α is set to 0.5. The parameter β is Eq.(4) is

set to 0.1. To reduce the computation complexity in Eq.(4), we use a hierarchical approach with three levels pyramid. The displacements resolution of each voxel on S are $(-4, -2, 0, 2, 4)$ mm, $(-2, -1, 0, 1, 2)$ mm and $(-1, -0.5, 0, 0.5, 1)$ mm along each dimension from $3th$ pyramid level to $1th$ pyramid level. The number of labels at each level is $5^3 = 125$ for 3D image. It takes between 30 to 100 seconds for most of the examples.

Evaluation Strategy. In order to assess the performance of the method, both simulated data and clinical 4D CT patients' data were used. To generate synthesized test datasets, clinical CT images with no artifacts were each divided into two sub-images partially overlapping each other. Then, known motion deformation vectors were applied to one sub-image to produce the corresponding moving image. Some 3D feature points were identified as the landmarks. The landmarks distance errors (LDE) between the resulting artifact-reduced images and the original images were computed as the metric. Our experiments studied the following aspects of the method: (1)the average and standard deviation of LDE; (2) the sensitivity to the initial registration methods; and (3) the sensitivity of the flow propagation. Besides the simulation evaluation, the results on clinical 4D CT images with artifacts were compared to those obtained by the commercial software.

Data Collection. For the synthesized datasets, five lung 3D CT images without artifacts were used. Each CT image consists of 40 slices with a resolution of 0.98mm×0.98mm×2mm. While dividing the 3D CT image into two sub-images, we set the overlapping between the two sub-images to be 10 slices, which indicates that there were 20 mm displacements along z-dimension between the fixed and the moving images. The number of landmarks identified for the measure of LDE was seven in each 3D CT image.

For the clinical test datasets, we use the images acquired by a 40-slice multidetector CT scanner (Siemens Biograph) operating in helical mode. The amplitude of the respiratory motion was monitored using a strain belt with a pressure sensor (Anzai, Tokyo, Japan). The respiratory phase at each time point was computed by the scanner console software via renormalization of each respiratory period by the period-specific maxima and minima. The Siemens Biograph 40 software was used to sort raw 4D CT images retrospectively into respiratory phase-based bins of phase-specific 3D CT images.

3.2 Results and Discussion

The results are summarized in Table 1 showing the average (avg) and standard deviation (std) of LDE's. We show the LDE's for each of the five synthesized datasets (1) while no registration operation was applied, i.e., simply stacking the two sub-images; (2) after applying the initial registration; and (3) after applying the proposed method. In our method, the combined affine and B-Spline registration were used as the initial registration method. We used the elastix tools [16] in our experiments. From Table 1, we can see that the LDE's were significantly reduced after initial registration. While after applying our method,

Table 1. Landmarks distances errors (LDE) (avg±std mm)

Method	Data 1	Data 2	Data 3	Data 4	Data 5	Avg
Before registration	24.8±2.6	26.4±2.8	25.1±2.3	27.8±3.8	26.1±1.2	26.1±2.4
After initial registration (Affine + B-Spline)	1.9±1.0	2.7±1.5	2.6±1.7	3.4±1.8	2.8±0.8	2.7±1.9
After the proposed method	1.0±0.2	1.3±0.9	1.5±1.4	1.4±1.1	2.6±1.2	1.5±0.9
Initial Registration Method						
Affine	1.6±0.7	2.3±1.7	2.1±1.4	2.4±1.5	2.8±1.5	2.2±1.4
B-Spline	1.0±0.2	1.9±1.4	1.7±1.5	1.5±0.9	2.7±1.4	1.8±1.0

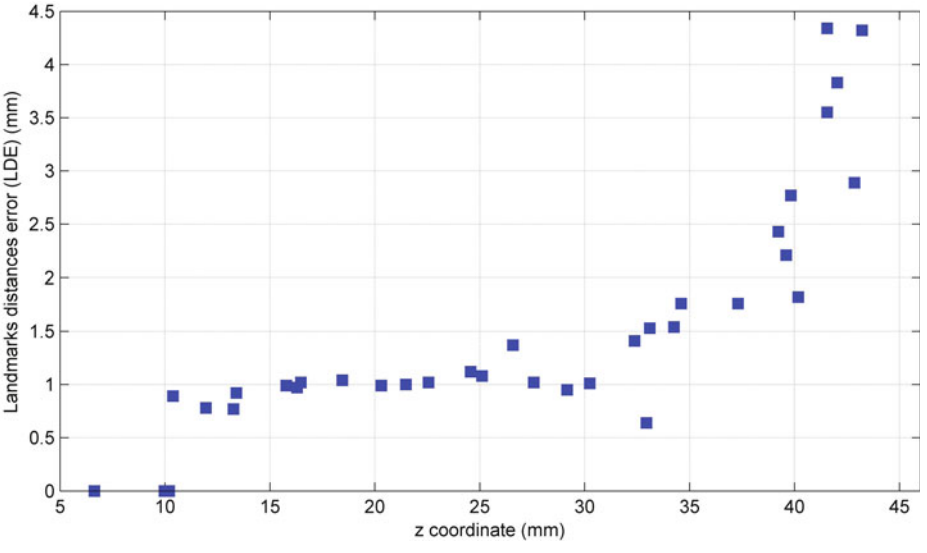


Fig. 4. LDE with respect to the z coordinate

the LDE's were further decreased by 42% from 2.7 mm to 1.5 mm on average. The standard deviation was also further decreased from 1.9 mm to 0.9 mm. To evaluate the sensitivity to the initial registration methods, we used the affine and the B-Spline as the initial registration in our method separately. The LDE's are shown in the last two rows of Table 1. The B-Spline initial registration achieved better results that the affine registration. The combined affine and B-Spline registration gave the best results. Overall, the differences were not significant. Thus, we conclude that our method is not sensitive to the initial registration with an error up to 4 mm because the search space in graph cuts of our algorithm is limited to 4 mm, which is shown in parameters setting section.

The flow estimation in the non-overlap region of the moving image is challenging due to the non-rigid deformation. Since the seam flow is propagated by solving a Laplace Equation, the LDE should increase as the distance of the

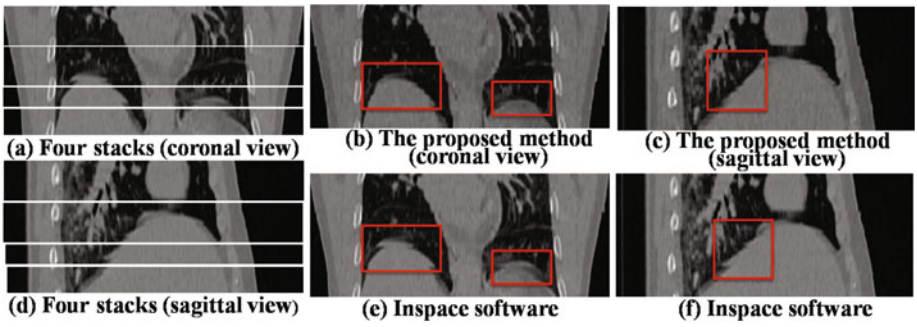


Fig. 5. The comparison of the proposed method with the commercial 4D CT software

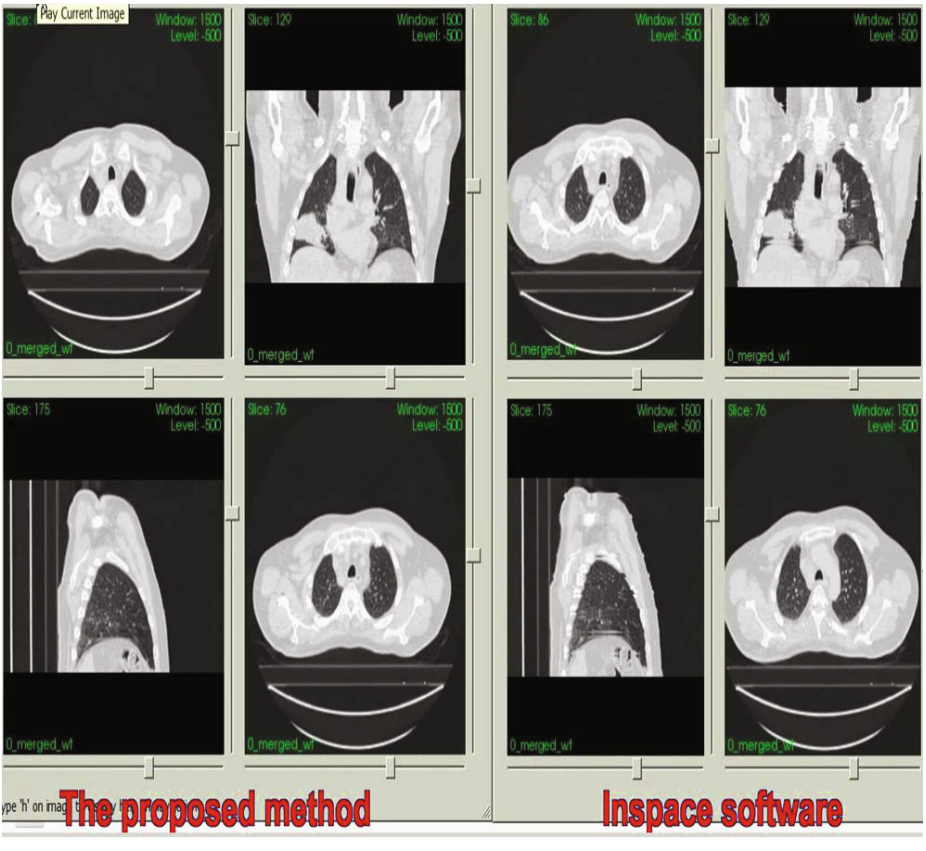


Fig. 6. More example comparison results with the commercial 4D CT software

landmark from the interface seam increases. To analyze the propagation behaviors of the seam flow in the moving image, we plotted the LDE's of all the landmarks based on their z-coordinates (note that a large z-coordinate indicates that the landmark is far away from the interface seam for stitching). Figure 4 shows that the LDE's were less than 2 mm when the z- coordinates of the landmarks were smaller than 40 mm (i.e., 20 slices in our data). The largest LDE observed in our experiment was about 4.4 mm. For the clinical 4D CT images, no method can guarantee the computed deformation field is correct in the non-overlap region of the moving image, especially, when the non-overlap region is large. Fortunately, one image stack commonly contains about 20 slices in our clinical helical 4D CT images. Thus, our results indicate that the method is stable in the clinical setting.

For the 4D CT images acquired in the helical mode, to the best of our knowledge, there are no known algorithms designated to reduce the reconstruction artifacts. Thus, comparing with the commercial 4D CT software (Inspace software) is our best choice. In the two input image stacks, the one in a better breath period was chosen as the fixed image and the other as the moving image. The quality of the resulting image was evaluated by the three medical experts. All observers identified much fewer artifacts in the images produced by the proposed method than those output by the Inspace. Example results are shown in Figure 5 and Figure 6.

4 Conclusion

An effective and simple method for reducing the magnitude of artifacts in helical 4D CT images was presented. The concept of seam flow was introduced to solve the misalignment problem. The presented method was evaluated on simulated data with promising performance. The results on clinical 4D CT images were compared to the commercial software and all medical experts identified fewer artifacts in the resulting images obtained by the proposed method than those by the commercial software. In conclusion, the reported approach is promising to improve the quality of 4D CT image and to reduce the artifacts directly from the reconstructed images.

Acknowledgement

This research was supported in part by the NSF grants CCF-0830402 and CCF-0844765, and the NIH grants R01 EB004640 and K25 CA123112.

References

1. Low, D., Nystrom, M., Kalinin, E., et al.: A method for the reconstruction of four-dimensional synchronized CT scans acquired during free breathing. Medical Physics 30, 1254–1263 (2003)
2. Rietzel, E., Pan, T., Chen, G.: Four-dimensional computed tomography: Image formation and clinical protocol. Medical Physics 32, 974–989 (2005)

3. Yamamoto, T., Langner, U., Loo, B.W., et al.: Retrospective analysis of artifacts in four-dimensional CT images of 50 abdominal and thoracic radiotherapy patients. International Journal of Radiation Oncology 72, 1250–1258 (2008)
4. Han, D., Byouth, J., Wu, X., et al.: Characterization and identification of spatial artifacts during 4D CT imaging. In: AAPM 2010 (2010)
5. Kwatra, V., Schödl, A., Essa, I., et al.: Graphcut textures: image and video synthesis using graph cuts. ACM Transactions on Graphics 22, 277–286 (2003)
6. Jia, J., Tang, C.K.: Image stitching using structure deformation. PAMI 30, 617–631 (2008)
7. Ehrhardt, J., Werner, R., Frenzel, T., Lu, W., Low, D., Handels, H.: Analysis of free breathing motion using artifact reduced 4D CT image data (2007)
8. McClelland, J., Blackall, J., Tarte, S., Chandler, A., Hughes, S., Ahmad, S., Landau, D., Hawkes, D.: A continuous 4D motion model from multiple respiratory cycles for use in lung radiotherapy. Medical Physics 33, 3348 (2006)
9. Li, K., Wu, X., Chen, D.Z., Sonka, M.: Optimal surface segmentation in volumetric images-a graph-theoretic approach. PAMI 28, 119–134 (2006)
10. Song, Q., Wu, X., Liu, Y., Smith, M., Buatti, J., Sonka, M.: Optimal graph search segmentation using arc-weighted graph for simultaneous surface detection of bladder and prostate. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 827–835. Springer, Heidelberg (2009)
11. Han, D., Sonka, M., Bayouth, J., Wu, X.: Optimal multiple-seams search for image resizing with smoothness and shape prior. The Visual Computer 26, 749–759 (2010)
12. Han, D., Wu, X., Sonka, M.: Optimal multiple surfaces searching for video/image resizing-a graph-theoretic approach. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 1026–1033 (2009)
13. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. PAMI 23, 1222–1239 (2001)
14. Pritch, Y., Kav-Venaki, E., Peleg, S.: Shift-map image editing. In: ICCV 2009, Kyoto (2009)
15. Pérez, P., Gangnet, M., Blake, A.: Poisson image editing. ACM Transactions on Graphics 22, 313–318 (2003)
16. Klein, S., Staring, M., Murphy, K., Viergever, M., Pluim, J.: Elastix: a toolbox for intensity-based medical image registration. IEEE Transactions on Medical Imaging 29 (2010)

Comparative Validation of Graphical Models for Learning Tumor Segmentations from Noisy Manual Annotations

Frederik O. Kaster[1,2], Bjoern H. Menze[3,4], Marc-André Weber[5], and Fred A. Hamprecht[1]

[1] Heidelberg Collaboratory for Image Processing, University of Heidelberg, Germany
frederik.kaster@iwr.uni-heidelberg.de,
fred.hamprecht@iwr.uni-heidelberg.de
[2] German Cancer Research Center, Heidelberg, Germany
[3] CSAIL, Massachusetts Institute of Technology, Cambridge MA, USA
menze@csail.mit.edu
[4] INRIA Sophia-Antipolis Méditerrannée, France
[5] Department of Diagnostic Radiology, University of Heidelberg, Germany
MarcAndre.Weber@med.uni-heidelberg.de

Abstract. Classification-based approaches for segmenting medical images commonly suffer from missing ground truth: often one has to resort to manual labelings by human experts, which may show considerable intra-rater and inter-rater variability. We experimentally evaluate several latent class and latent score models for tumor classification based on manual segmentations of different quality, using approximate variational techniques for inference. For the first time, we also study models that make use of image feature information on this specific task. Additionally, we analyze the outcome of hybrid techniques formed by combining aspects of different models. Benchmarking results on simulated MR images of brain tumors are presented: while simple baseline techniques already gave very competitive performance, significant improvements could be made by explicitly accounting for rater quality. Furthermore, we point out the transfer of these models to the task of fusing manual tumor segmentations derived from different imaging modalities on real-world data.

1 Introduction and Related Work

The use of machine learning methods for computer-assisted radiological diagnostics faces a common problem: In most situations, it is impossible to obtain reliable ground-truth information for e.g. the location of a tumor in the images. Instead one has to resort to manual segmentations by human labelers, which are necessarily imperfect due to two reasons. Firstly, humans make labeling mistakes due to insufficient knowledge or lack of time. Secondly, the medical images upon which they base their judgment may not have sufficient contrast to discriminate between tumor and non-tumor tissue. In general, this causes both a systematic

B. Menze et al. (Eds.): MICCAI 2010 Workshop MCV, LNCS 6533, pp. 74–85, 2011.

bias (tumor outlines are consistently too large or small) and a stochastic fluctuation of the manual segmentations, both of which depend on the specific labeler and the specific imaging modality.

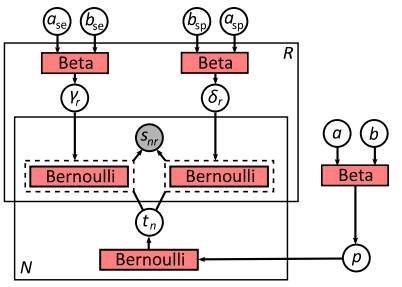
One can alleviate this problem by explicitly modelling the decision process of the human raters: in medical image analysis, this line of research started with the STAPLE algorithm (Warfield et al., 2004) and its extensions (Warfield et al., 2008), while in the field of general computer vision, it can already be traced back to the work of Smyth et al. (1995). Similar models were developed in other application areas of machine learning [Raykar et al. (2010), Whitehill et al. (2009), Rogers et al. (2010)]: some of them make also use of image information and produce a classifier, which may be applied to images for which no annotations are available. The effect of the different imaging modalities on the segmentation has not yet found as much attention.

In this paper, we systematically evaluated these competing methods as well as novel hybrid models for the task of computer-assisted tumor segmentation in radiological images: we used the same machinery on annotations provided by multiple human labelers with different quality and on annotations based on multiple imaging modalities. While traditionally these methods have been tackled by expectation maximization (EM; Dempster et al., 1977), we formulate the underlying inference problems as probabilistic graphical models (Koller and Friedman, 2009) and thereby render them amenable to generic inference methods (see Fig. 1). This facilitates the inference process and makes it easier to study the effect of modifications on the final inference results.

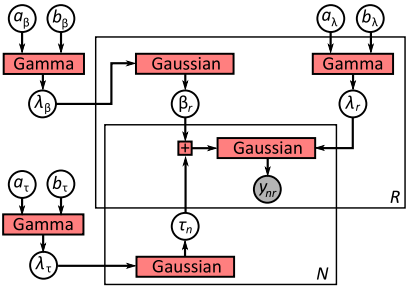
2 Theory and Modelling

Previous models. In the following we detail on earlier and novel probabilistic models studied in the present work. In the STAPLE model proposed by Warfield (2004, Fig. 1(a)) the discrete observations $s_{nr} \in \{0, 1\}$ are noisy views on the true scores $t_n \in \{0, 1\}$, with $n \in \{1, \ldots, N\}$ indexing the image pixels and $r \in \{1, \ldots, R\}$ indexing the raters. The r-th rater is characterized by the sensitivity γ_r and the specificity $1-\delta_r$, and the observation model is $s_{nr} \sim t_n \mathrm{Ber}(\gamma_r) + (1 - t_n)\mathrm{Ber}(\delta_r)$, with "Ber" denoting a Bernoulli distribution. A Bernoulli prior is given for the true class: $t_n \sim \mathrm{Ber}(p)$. While the original formulation fixes $p = 0.5$ and uses uniform priors for γ_r and δ_r, we modify the priors to fulfil the conjugacy requirements for the chosen variational inference techniques: hence we impose beta priors on $\gamma_r \sim \mathrm{Beta}(a_{\mathrm{se}}, b_{\mathrm{se}})$, $\delta_r \sim \mathrm{Beta}(b_{\mathrm{sp}}, a_{\mathrm{sp}})$ and $p \sim \mathrm{Beta}(a_{\mathrm{p}}, b_{\mathrm{p}})$. The latter distribution is introduced in order to learn the share of tumor tissue among all voxels from the data.

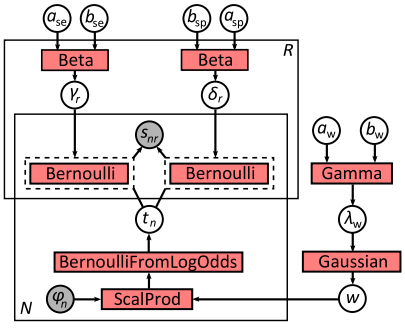
The model by Raykar et al. (2010, Fig. 1(c)) is the same as (Warfield et al., 2004) except for the prior on t_n: the authors now assume that a feature vector ϕ_n is observed at the n-th pixel and that $t_n \sim \mathrm{Ber}\big(\{1+\exp(-w^\top \phi_n)\}^{-1}\big)$ follows a logistic regression model. A Gaussian prior is imposed on $w \sim \mathcal{N}(0, \lambda_{\mathrm{w}}^{-1} I)$. In contrast to (Warfield et al., 2004), they obtain a classifier that can be used to predict the tumor probability on unseen test images, for which one has access



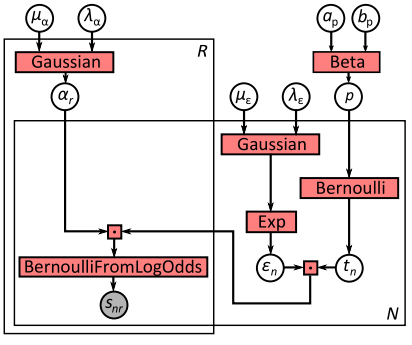
(a) Warfield et al. (2004)



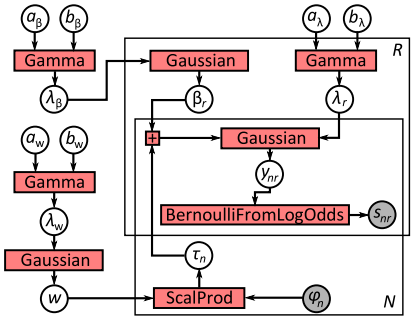
(b) Warfield et al. (2008)



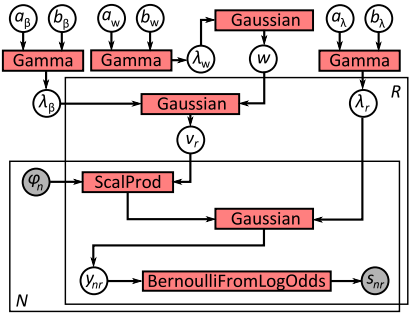
(c) Raykar et al. (2010)



(d) Whitehill et al. (2009)



(e) Hybrid model 2



(f) Hybrid model 4

Fig. 1. Graphical model representations. Red boxes correspond to factors, circles correspond to observed (gray) and unobserved (white) variables. Solid black rectangles are plates indicating an indexed array of variables (Buntine, 1994). The dashed rectangles are *"gates"* denoting a mixture model with a hidden selector variable (Minka and Winn, 2009).

to the features ϕ_n but not to the annotations s_{nr}. One may hypothesize that the additional information of the features ϕ_n can help to resolve conflicts: in a two-rater scenario, one can decide that the rater has less noise who labels pixels with similar ϕ_n more consistently. In our graphical model formulation, we add a gamma prior for the weight precision $\lambda_w \sim \text{Gam}(a_w, b_w)$.

Whitehill et al. (2009, Fig. 1(d)) propose a model in which the misclassification probability depends on both the pixel and the rater: $s_{nr} \sim \text{Ber}\big(\{1 + \exp(-t_n \alpha_r \epsilon_n)\}^{-1}\big)$ with the rater accuracy $\alpha_r \sim \mathcal{N}(\mu_\alpha, \lambda_\alpha^{-1})$ and the pixel difficulty ϵ_n with $\log(\epsilon_n) \sim \mathcal{N}(\mu_\epsilon, \lambda_\epsilon^{-1})$ (this parameterization is chosen to constrain ϵ_n to be positive).

In the continuous variant of STAPLE by Warfield et al. (2008, Fig. 1(b)), the observations y_{nr} are continuous views on a continuous latent score τ_n. The r-th rater can be characterized by a bias β_r and a noise precision λ_r: $y_{nr} \sim \mathcal{N}(\tau_n + \beta_r, \lambda_r^{-1})$, with a Gaussian prior on the true scores: $\tau_n \sim \mathcal{N}(0, \lambda_\tau^{-1})$. In contrast to the original formulation, we add Gaussian priors on the biases, i.e. $\beta_r \sim \mathcal{N}(0, \lambda_\beta^{-1})$. For the precisions of the Gaussians, we use gamma priors: $\lambda_\tau \sim \text{Gam}(a_\tau, b_\tau)$, $\lambda_\beta \sim \text{Gam}(a_\beta, b_\beta)$ and $\lambda_r \sim \text{Gam}(a_\lambda, b_\lambda)$. Note that when thresholding the continuous scores, the tumor boundary may shift because of the noise, but misclassifications far away from the boundary are unlikely: this is an alternative to (Whitehill et al., 2009) for achieving a non-uniform noise model.

Novel hybrid models. We also study four novel hybrid models, which incorporate all aspects of the previous proposals simultaneously: while they provide a classifier as in (Raykar et al., 2010), they do not assume misclassifications to occur everywhere equally likely. In the simplest variant (hybrid model 1), we modify the model from (Warfield et al., 2008) by a linear regression model for $\tau_n \sim \mathcal{N}(w^\top \phi_n, \lambda_\tau^{-1})$ with $w \sim \mathcal{N}(0, \lambda_\text{w}^{-1})$. Note that this model predicts a (noisy) linear relationship between the distance transform values y_{nr} and the features ϕ_n, while experimentally the local image appearance saturates in the interior of the tumor or the healthy tissue. To alleviate this concern (hybrid model 2, Fig. 1(e)), one can interpret y_{nr} as an unobserved malignancy score, which influences the (observed) binary segmentations s_{nr} via $s_{nr} \sim \text{Ber}\big(\{1 + \exp(-y_{nr})\}^{-1}\big)$. This is a simplified version of the procedure presented in (Rogers et al., 2010), with a linear regression model for the latent score instead of a Gaussian process regression. Alternatively one can model the raters as using a biased weight vector rather than having a biased view on an ideal score, i.e. $y_{rn} \sim \mathcal{N}(v_r^\top \phi_n, \lambda_r^{-1})$ with $v_r \sim \mathcal{N}(w, \lambda_\beta^{-1} I)$. Again the score y_{nr} may be observed directly as a distance transform (hybrid model 3) or indirectly via s_{nr} (hybrid model 4, Fig. 1(f)).

Inference. For the graphical models considered here, exact inference by the junction tree algorithm is infeasible owing to the high number of variables and the high number of V structures, which lead to a nearly complete graph after moralization (Koller and Friedman, 2009). However, one can perform approximate inference using e.g. variational message passing (Winn and Bishop, 2005): the true posterior for the latent variables is approximated by the closest factorizing distribution (as measured by the Kullback-Leibler distance), for which inference is tractable. As a prerequisite, all priors must be conjugate; this holds for all models discussed above except (Whitehill et al., 2009). Here we cannot apply the generic variational message passing scheme to this model, and show the results from the EM inference algorithm provided by the authors instead.

We employed the INFER.NET 2.3 Beta implementation for variational message passing (Minka et al., 2009) to perform inference on the algorithms by

Warfield et al. (2004), Warfield et al. (2008), Raykar et al. (2010) and the four hybrid models. The default value of 50 iteration steps was found to be sufficient for convergence, since doubling the number of steps led to virtually indistinguishable results. For the algorithm by Whitehill et al. (2009), we used the GLAD 1.0.2 reference implementation[1]. Alternative choices for the generic inference method would have been expectation propagation (Minka, 2001) and Gibbs sampling (Gelfand and Smith, 1990). We experimentally found out that expectation propagation had considerably higher memory requirements than variational message passing for our problems, which prevented its use for our problems on the available hardware. Gibbs sampling was not employed since some of the factors incorporated in our models (namely gates and factor arrays) are not supported by the current INFER.NET implementation.

We also compared against three baseline procedures: majority voting, training a logistic regression classifier from the segmentations of every single rater and averaging the classifier predictions (ALR), and training a logistic regression classifier on soft labels (LRS): if S out of R raters voted for tumor in a certain pixel, it was assigned the soft label $S/R \in [0, 1]$.

3 Experiments

We performed two experiments in order to study the influences of labeler quality and imaging modality separately. In the first experiment, we collected and fused multiple human annotations of varying quality based on one single imaging modality: here we used simulated brain tumor measurements for which ground truth information about the true tumor extent was available, so that the results could be evaluated quantitatively. In the second experiment, we collected and fused multiple human annotations, which were all of high quality but had been derived from different imaging modalities showing similar physical changes caused by glioma infiltration with different sensitivity.

Human raters. Simulated brain tumor MR images were generated by means of the TumorSim 1.0 software by Prastawa et al. (2009)[2]. The advantage of these simulations was the existence of ground truth about the true tumor extent (in form of probability maps for the distribution of white matter, gray matter, cerebrospinal fluid, tumor and edema). Our task was to discriminate between "pathological tissue" (tumor and edema) and "healthy tissue" (the rest). We used nine volumes: three for each tumor class that can be simulated by this software (ring-enhancing, uniformly enhancing and non-enhancing). Each volumetric images contained $256 \times 256 \times 181$ voxels and the three different imaging modalities (T_1-weighted with and without gadolinium enhancement and T_2-weighted) were considered perfectly registered with respect to each other. The feature vectors ϕ_i consisted of four features for each modality: gray value, gradient magnitude and the responses of a minimum and maximum filter within a 3×3 neighborhood. A

[1] `http://mplab.ucsd.edu/~jake/OptimalLabelingRelease1.0.2.tar.gz`
[2] `http://www.sci.utah.edu/releases/tumorsim_v1.0/TumorSim_1.0_linux64.zip`

row with the constant value 1 was added to learn a constant offset for the linear or logistic models (since there was no reason to assume that features values at the tumor boundary are orthogonal to the final weight vector).

The image volumes were segmented manually based on hypointensities in the T_1-weighted images, using the manual segmentation functionality of the ITK-SNAP 2.0 software[3]. In order to control the rater precision, time limits of 60, 90, 120 and 180 seconds for labeling a 3D volume were imposed and five segmentations were created for each limit: we expect the segmentations to be precise for generous time limits, and to be noisy when the rater had to label very fast. The set of raters was the same for the different time constraints, and the other experimental conditions were also kept constant across the different time constraints. This was statistically validated: the area under curve value of the receiver operating characteristic of the ground-truth probability maps compared against the manual segmentations showed a significant positive trend with respect to the available time ($p = 1.8 \times 10^{-4}$, F test for a linear regression model). Since tight time constraints are typical for the clinical routine, we consider this setting as realistic, although it does not account for rater bias.

We extracted the slices with the highest amount of tumor lesion, and partitioned them into nine data subsets in order to estimate the variance of segmentation quality measures, with each subset containing one third of the slices extracted from three different tumor datasets (one for each enhancement type). For memory reasons, the pixels labeled as "background" by all raters were randomly subsampled to reduce the sample size. A cross-validation scheme was used to test the linear and log-linear classifiers[4] on features ϕ_n not seen during the training process: we repeated the training and testing nine times and chose each of the data subsets in turn as the training dataset (and two different subsets as the test data).

The following default values for the hyperparameters were used: $a_{\mathrm{Se}} = 10$, $b_{\mathrm{Se}} = 2$, $a_{\mathrm{Sp}} = 10$, $b_{\mathrm{Sp}} = 2$, $a_{\mathrm{w}} = 2$, $b_{\mathrm{w}} = 1$, $a_{\mathrm{p}} = 2$, $b_{\mathrm{p}} = 2$, $a_\tau = 2$, $b_\tau = 1$, $a_\beta = 2$, $b_\beta = 1$, $a_\lambda = 2$, $b_\lambda = 1$. We confirmed in additional experiments that inference results changed only negligibly when these hyperparameters were varied over the range of a decade. In order to check the effect of the additional priors that we introduced into the models by Warfield et al. (2004), Warfield et al. (2008) and Raykar et al. (2010), we also ran experiments with exactly the same models as in the original papers (by fixing the corresponding variables or using uniform priors). However, this led to uniformly worse inference results than in our model formulations.

Multiple modalities. For evaluation on real-world measurements, we used a set of twelve multimodal MR volumes acquired from glioma patients (T_1-, T_2-, FLAIR- and post-gadolinium T_1-weighting), which had been affinely registered to the FLAIR volume: we used a automated multi-resolution mutual information registration procedure as included in the MedINRIA[5] software. Manual

[3] `http://www.itksnap.org/pmwiki/pmwiki.php?n=Main.Downloads`

[4] All except (Warfield et al., 2004), (Warfield et al., 2008), (Whitehill et al., 2009).

[5] `https://gforge.inria.fr/projects/medinria`

segmentations of pathological tissue (tumor and edema) were provided separately for every modality on 60 slices extracted from these volumes (20 axial, sagittal and coronal slices each of which intersecting with the tumor center). In these experiments, we propose to use the described models to infer a single probability map summarizing all tumor-induced changes in the different imaging modalities. In particular, we identify every modality as a separate "rater" with a specific and consistent bias with respect to the joint probability map inferred.

4 Results

Multiple raters. We studied several scenarios, i.e. several compositions of the rating committee. Here we exemplarily report the results for two of them: one with a majority of good raters (120/120/90, i.e. two raters with a 120 sec constraint and one rater with a 90 sec constraint) and one with a majority of poor raters (60/60/60/180/180, i.e. three raters with a 60 sec constraint, and two raters with a 180 sec constraint). Tables 1 and 2 show the results of various evaluation statistics both for training data (for which the human annotations were used) and test data. Sensitivity, specificity, correct classification rate (CCR) and Dice

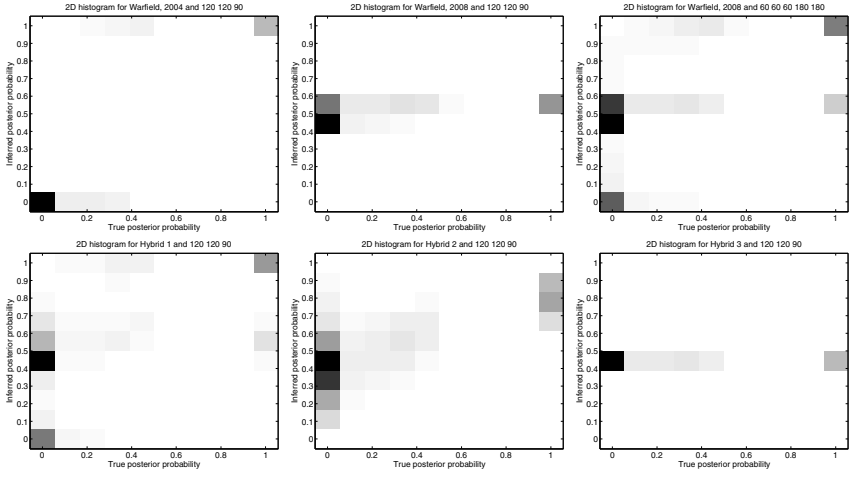


Fig. 2. Comparison of ground-truth (abscissa) and inferred posterior (ordinate) tumor probabilities, visualized as normalized 2D histograms. All histograms are normalized such that empty bins are white, and the most populated bin is drawn black. We show the inference results of (Warfield et al., 2004), (Warfield et al., 2008), and the hybrid models 1–3. The results of hybrid model 4 were similar to hybrid model 3, and the results of (Raykar et al., 2010) and (Whitehill et al., 2009) were similar to (Warfield et al., 2004). Mostly the two scenarios 120/120/90 and 60/60/60/180/180 gave similar results so that we show only the results for the former, with the exception of (Warfield et al., 2008) (top middle and top right). For the ideal inference method, all bins outside the main diagonal would be white; (Warfield et al., 2004) comes closest.

Table 1. Evaluation statistics for the training data (i.e. the manual annotations of the raters were used for inference), under the 120/120/90 scenario. The first three rows show the outcome of the three baseline techniques. The best result in each column is marked *by italics*, while **bold figures** indicate a significant improvement over the best baseline technique ($P < .05$, rank-sum test with multiple-comparison adjustment). Estimated standard deviations are given in parentheses. The outcome of the other scenarios was qualitatively similar (especially concerning the relative ranking between different inference methods). ALR = Averaged logistic regression. LRS = Logistic regression with soft labels. CCR = Correct classification rate (percentage of correctly classified pixels). AUC = Area Under Curve of the receiver operating characteristics curve obtained when thresholding the ground-truth probability map at 0.5. Dice = Dice coefficient of the segmentations obtained when thresholding both the inferred and the ground-truth probability map at 0.5.

	Specificity	Sensitivity	CCR	AUC	Dice
Majority vote	.987(007)	.882(051)	.910(032)	.972(008)	.827(020)
ALR	.953(018)	*.920(036)*	*.931(025)*	.981(005)	*.855(031)*
LRS	.953(019)	.919(037)	*.931(025)*	.981(005)	*.855(030)*
(Warfield et al., 2004)	.987(007)	.882(051)	.910(032)	.972(008)	.827(020)
(Warfield et al., 2008)	***1.000(001)***	.617(130)	.692(139)	**.989(003)**	.584(211)
(Raykar et al., 2010)	.988(006)	.886(045)	.913(028)	***.993(003)***	.830(024)
(Whitehill et al., 2009)	.988(004)	.913(016)	*.931(008)*	.980(003)	.845(063)
Hybrid model 1	.940(078)	.692(060)	.751(070)	.902(117)	.603(191)
Hybrid model 2	.972(019)	.716(048)	.770(057)	.953(015)	.628(163)

coefficient are computed from the binary images that are obtained by thresholding both the ground-truth probability map and the inferred posterior probability map at 0.5. If n_{fb} denotes the number of pixels that are thereby classified as foreground (tumor) in the ground truth and as background in the posterior probability map (and n_{bb}, n_{bf} and n_{ff} are defined likewise), these statistics are computed as follows:

$$\text{Sensitivity} = \frac{n_{\mathrm{ff}}}{n_{\mathrm{fb}} + n_{\mathrm{ff}}}, \qquad \text{Specificity} = \frac{n_{\mathrm{bb}}}{n_{\mathrm{fb}} + n_{\mathrm{bb}}},$$

$$\text{CCR} = \frac{n_{\mathrm{ff}} + n_{\mathrm{bb}}}{n_{\mathrm{ff}} + n_{\mathrm{bb}} + n_{\mathrm{bf}} + n_{\mathrm{fb}}}, \qquad \text{Dice} = \frac{2n_{\mathrm{ff}}}{2n_{\mathrm{ff}} + n_{\mathrm{bf}} + n_{\mathrm{fb}}}$$

Additionally we report the Area Under Curve (AUC) value for the receiver operating curve obtained by binarizing the ground-truth probabilities with a fixed threshold of 0.5 and plotting sensitivity against 1 − specificity while the threshold for the posterior probability map is swept from 0 to 1. Most methods achieved Dice coefficients in the range of 0.8–0.85, except for the models operating on a continuous score (the hybrid models and the model by Warfield et al. (2008)). Since our features were highly discriminative, even simple label fusion schemes such as majority voting gave highly competitive results. Qualitatively, there is little difference between these two scenarios (and the other ones under study).

Table 2. Evaluation statistics for the test data (i.e. the manual annotations of the raters were not used for inference), under the 120/120/90 scenario. Note that one can only employ the inference methods which make use of the image features ϕ_n and estimate a weight vector w: the unobserved test data labels are then treated as missing values and are marginalized over. To these examples we cannot apply the methods which only use the manual annotations: majority voting, (Warfield et al., 2004) and (Warfield et al., 2008). The results for the other scenarios were qualitatively similar (especially concerning the relative ranking between different inference methods). Cf. the caption of table 1 for further details.

	Sensitivity	Specificity	CCR	AUC	Dice
ALR	.937(017)	.924(038)	.928(029)	*.978(009)*	.837(065)
LRS	.936(017)	.925(038)	.928(029)	*.978(009)*	.837(066)
(Raykar et al, 2010)	.927(019)	***.937(031)***	*.936(025)*	.977(013)	*.853(038)*
Hybrid model 1	.851(152)	.735(181)	.760(167)	.852(172)	.619(142)
Hybrid model 2	***.973(013)***	.727(174)	.786(116)	.952(026)	.667(084)

Some graphical models perform better than the baseline methods on the training data, namely (Raykar et al., 2010) and (Warfield et al., 2008). However, they bring no improvement on the test data.

Unexpectedly, the hybrid models perform worse and with lesser stability than the simple graphical models, and for hybrid models 3 and 4, the inference converges to a noninformative posterior probability of 0.5 everywhere. It should be noted that the posterior estimates of the rater properties did not differ considerably between corresponding algorithms such as (Warfield et al., 2008) and (Raykar et al., 2010), hence the usage of image features does not allow one to distinguish between better and poorer raters more robustly.

In order to account for partial volume effects and blurred boundaries between tumor and healthy tissue, it is preferable to visualize the tumors as soft probability maps rather than as crisp segmentations. In Fig. 2, we compare the ground-truth tumor probabilities with the posterior probabilities following from the different models. Some models assume a latent binary class label, namely (Warfield et al., 2004), (Raykar et al., 2010) and (Whitehill et al., 2009): they tend to sharpen the boundaries between tumor and healthy tissue overly, while the latent score models (all others) smooth them. One can again note that the true and inferred probabilities are completely uncorrelated for hybrid model 3 (and 4).

Multiple modalities. The optimal delineation of tumor borders in multi-modal image sequences and obtaining ground truth remains difficult. So, in the present study we confine ourselves to a first, qualitative comparison of the different models. Fig. 3 shows the posterior probability maps for a real-world brain image example. The results of (Warfield et al., 2004) and (Warfield et al., 2008) can be regarded as extreme cases: the former yields a crisp segmentation without accounting for uncertainty near the tumor borders, while the latter assigns a probability near 0.5 to all pixels and is hence inappropriate for this task. Hybrid

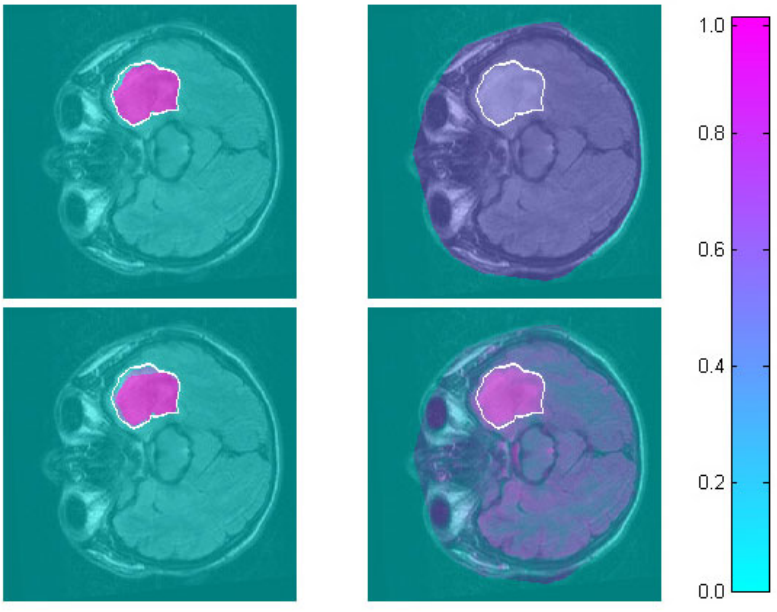


Fig. 3. Example of a FLAIR slice with manual segmentation of tumor drawn on the same FLAIR image (white contour), and inferred mean posterior tumor probability maps for (Warfield et al., 2004) (top left), (Warfield et al., 2008) (top right), (Whitehill et al., 2009) (bottom left) and hybrid model 2 (bottom right). The results of hybrid model 3 and 4 were nearly identical to (Warfield et al., 2008), the results of hybrid model 1 to model 2, and the results of (Raykar et al., 2010) to (Whitehill et al., 2009). Tumor probabilities outside the skull mask were automatically set to 0. We recommend to view this figure in the colored online version.

model 1 (or 2) and (Whitehill et al., 2009) or (Raykar et al., 2010) are better suited for the visualization of uncertainties.

5 Discussion and Outlook

In this study, we introduced graphical model formulations to the task of fusing noisy manual segmentations: e.g. the model by Raykar et al. (2010) had not been previously employed in this context, and it was found to improve upon simple logistic regression on the training data. However, these graphical models do not always have an advantage over simple baseline techniques: compare the results of (Warfield et al., 2004) to majority voting. Hybrid models combining the aspects of several models did not fare better than simple models. This ran contrary to our initial expectations, which were based on two assumptions: that different pixels have a different probability of being mislabeled, and that it is possible to detect these pixels based on the visual content (these pixels would be assigned high scores far away from the decision boundary). This may be an

artifact of our time-constrained labeling experiment: if misclassifications can be attributed mostly to chance or carelessness rather than to ignorance or visual ambiguity, these assumptions obviously do not hold, and a uniform noise model as in (Warfield et al., 2004) or (Raykar et al., 2010) should be used instead. It is furthermore not yet understood why the slight model change between hybrid models 1 / 2 and hybrid models 3 / 4 leads to the observed failure of inference. For the future, it should be checked if these effects arise from the use of an approximate inference engine or are inherent to these models: hence unbiased Gibbs sampling results should be obtained for comparison purposes, using e.g. the WinBUGS modelling environment (Lunn et al., 2000).

The use of simulated data for the main evaluation is the main limitation of our approach, as simulations always present a simplification of reality and cannot account for all artifacts and other causes for image ambiguity that are encountered in real-world data. However, this limitation is practically unavoidable, since we are assessing the imperfections of the currently best clinical practice for the precise delineation of brain tumors, namely manual segmentation of MR images by human experts. This assessment requires a superior gold standard by which the human annotations may be judged, and this can only be obtained from an *in silico* ground truth. For animal studies, a possible alternative lies in sacrificing the animals and delineating the tumor on histological slices which can be examined with better spatial resolution. However, these kinds of studies are costly and raise ethical concerns. Additionally, even expert pathologists often differ considerably in their assessment of histological images (Giannini et al., 2001).

Better segmentations could presumably be achieved by two extensions: More informative features could be obtained by registration of the patient images to a brain atlas, e.g. in the spirit of (Schmidt et al., 2005). An explicit spatial regularization could be achieved by adding an MRF prior on the latent labels or scores, and employing a mean-field approximation (Zhang, 1992) to jointly estimate the optimum segmentation and the model parameters by EM.

References

Buntine, W.: Operations for Learning with Graphical Models. Journal of Artificial Intelligence Research 2, 159–225 (1994)

Dempster, A., Laird, N., Rubin, D., et al.: Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society. Series B (Methodological) 39(1), 1–38 (1977)

Gelfand, A.E., Smith, A.F.: Sampling-Based Approaches to Calculating Marginal Densities. Journal of the American Statistical Association 85(410), 398–409 (1990)

Giannini, C., Scheithauer, B., Weaver, A., et al.: Oligodendrogliomas: reproducibility and prognostic value of histologic diagnosis and grading. Journal of Neuropathology & Experimental Neurology 60(3), 248 (2001)

Koller, D., Friedman, N.: Probabilistic Graphical Models – Principles and Techniques. MIT Press, Cambridge (2009)

Lunn, D., Thomas, A., Best, N., et al.: WinBUGS – A Bayesian modelling framework: Concepts, structure and extensibility. Statistics and Computing 10, 325–337 (2000)

Minka, T.: Expectation Propagation for approximate Bayesian inference. In: Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence, pp. 362–369 (2001)

Minka, T., Winn, J., Guiver, J., et al.: Infer.NET 2.3. Microsoft Research, Cambridge (2009), http://research.microsoft.com/infernet

Minka, T., Winn, J.: Gates. In: Koller, D., Schuurmans, D., Bengio, Y., Bottou, L. (eds.) Advances in Neural Information Processing Systems 21, pp. 1073–1080. MIT Press, Cambridge (2009)

Prastawa, M., Bullitt, E., Gerig, G.: Simulation of Brain Tumors in MR Images for Evaluation of Segmentation Efficacy. Medical Image Analysis 13(2), 297–311 (2009)

Raykar, V.C., Yu, S., Zhao, L.H., et al.: Learning From Crowds. Journal of Machine Learning Research 11, 1297–1322 (2010)

Rogers, S., Girolami, M., Polajnar, T.: Semi-parametric analysis of multi-rater data. Statistics and Computing 20(3), 317–334 (2010)

Schmidt, M., Levner, I., Greiner, R., et al.: Segmenting Brain Tumors using Alignment-Based Features. In: Proceedings of the Fourth International Conference on Machine Learning and Applications (ICMLA), pp. 215–220 (2005)

Smyth, P., Fayyad, U., Burl, M., et al.: Inferring Ground Truth From Subjective Labelling of Venus Images. In: Tesauro, G., Toretzy, D., Leen, T. (eds.) Advances in Neural Information Processing Systems, vol. 7, pp. 1085–1092. MIT Press, Cambridge (1995)

Warfield, S., Zou, K., Wells, W.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Transactions on Medical Imaging 23(7), 903–921 (2004)

Warfield, S., Zou, K., Wells, W.: Validation of image segmentation by estimating rater bias and variance. Philosophical Transactions of the Royal Society A 366(1874), 2361–2375 (2008)

Whitehill, J., Ruvolo, P.: fan Wu, T., et al.: Whose Vote Should Count More: Optimal Integration of Labels from Labelers of Unknown Expertise. In: Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C.K.I., Culotta, A. (eds.) Advances in Neural Information Processing Systems 22, pp. 2035–2043. MIT Press, Cambridge (2009)

Winn, J., Bishop, C.: Variational Message Passing. Journal of Machine Learning Research 6, 661–694 (2005)

Zhang, J.: The mean field theory in EM procedures for Markov random fields. IEEE Transactions on Signal Processing 40(10), 2570–2583 (1992)

Localization of 3D Anatomical Structures Using Random Forests and Discrete Optimization*

René Donner[1,2], Erich Birngruber[1], Helmut Steiner[1], Horst Bischof[2], and Georg Langs[3]

[1] Computational Image Analysis and Radiology Lab, Department of Radiology, Medical University of Vienna, Austria
rene.donner@meduniwien.ac.at
[2] Institute for Computer Graphics and Vision, Graz University of Technology, Austria
[3] Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, MA, USA

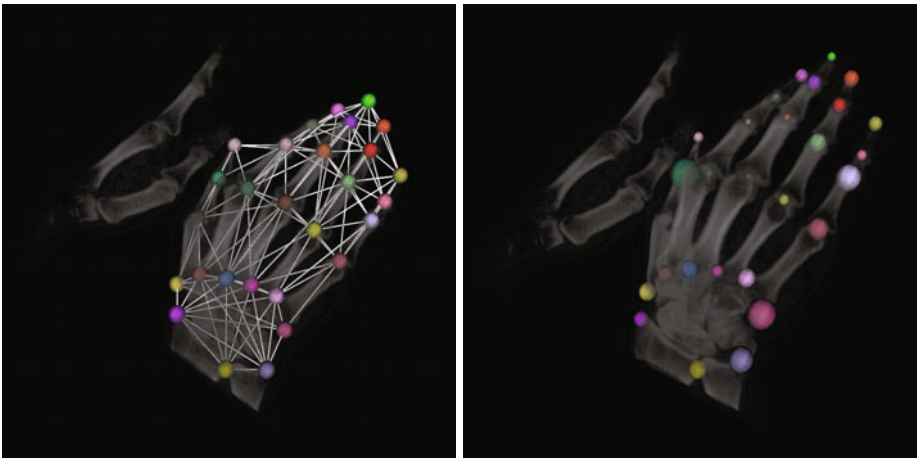
Abstract. In this paper we propose a method for the automatic localization of complex anatomical structures using interest points derived from Random Forests and matching based on discrete optimization. During training landmarks are annotated in a set of example volumes. A sparse elastic model encodes the geometric constraints of the landmarks. A Random Forest classifier learns the local appearance around the landmarks based on Haar-like 3D descriptors. During search we classify all voxels in the query volume. This yields probabilities for each voxel that indicate its correspondence with the landmarks. Mean-shift clustering obtains a subset of 3D interest points at the locations with the highest similarity in a local neighboorhood. We encode these points together with the conformity of the connecting edges to the learnt geometric model in a Markov Random Field. By solving the discrete optimization problem the most probable locations for each model landmark are found in the query volume. On a set of 8 hand CTs we show that this approach is able to consistently localize the model landmarks (finger tips, joints, etc) despite the complex and repetitive structure of the object.

1 Introduction

The reliable, fast segmentation of anatomical structures is a central issue in medical image analysis. It has been tackled by a number of powerful approaches. Among them are Active Shape Models / Active Appearance Models [5], Active Feature Models [12], Graph-Cuts [2], Active Contours [10], or Level-Set approaches [13]. All of these approaches require the correct localization of the

* This work was partly supported by the European Union FP7 Project KHRESMOI (FP7-257528), by the NSF IIS/CRCNS 0904625 grant, the NSF CAREER 0642971 grant, the NIH NCRR NAC P41-RR13218 and the NIH NIBIB NAMIC U54-EB005149 grant. Further supported by the Austrian National Bank grants COBAQUO (12537), BIOBONE (13468) and AORTAMOTION (13497).

B. Menze et al. (Eds.): MICCAI 2010 Workshop MCV, LNCS 6533, pp. 86–95, 2011.



(a) (b)

Fig. 1. (a) Employed data set including the manually annotated landmarks and the connectivities used to build the geometric model. (b) Depicts the localization result with the median residual distance between localization result and ground truth from all leave-one-out runs as radius for each sphere. Note the high accuracy of the localization.

initial model positions or seeds points within or close to the desired object. While most research has focused on the segmentation or analysis of individual regions of interest, the initialization was often performed manually or by application specific and often heuristic approaches. In this paper we propose a generic method that identifies anatomical structures in a global search framework. It learns the local appearance of landmarks, and an elastic shape constraint from annotated training volumes. During search a classifier is used for the generation of candidate points, and the final location is identified by discrete optimization.

The approach proposed in this paper is related to two lines of previous work:

1. Single object localization [18] have demonstrated an efficient voting scheme for localizing anatomical structures - in a sliding window technique each block predicts the location of the object based a priori knowledge learnt from blocks' appearances during a training phase. The result is a very robust estimate of the center of the object and its rotation, but no information is obtained about subparts of the object, although this could be accomplished by using a multi-scale approach narrowing in on the subparts. [6] presented a fast approach to localize individual organs in full-body CTs using 3D feature similar to Haar-wavelets with random offsets. Through these offsets they incorporate long range contextual information which allows to separate similar-looking objects to a certain extent, but the presented results did not comprise any small and repetitive structures. Random forests were used to classify the volumes, which through their parallel nature lend themselves nicely to a GPU-based implementation, resulting in very fast computation.

2. Localizing Complex Structures / Incorporating Subpart Interdependencies [14] parse full body CT data in a hierarchical fashion but are concerned with finding larger scale organs. They first search for one salient slice in each dimension and consequently only localize landmarks within these slices. While greatly speeding up the localization this only works for objects which are rather large in respect to the volume size, as all the objects have to be visible in at least one of the 3 slices. This assumption does not hold true, e. g., in case of the inclined main plain of a hand CT. [17] pose the problem as a dependency graph of individual localization steps, whose order and mutual influence is determined by information theoretic measures. The recently proposed work in [1] is most similar to ours, but randomly selects several candidate interest points within a region of high classification probability, while our mean-shift based approach will yield only one, more stable, interest point. [1] is additionally capable of dealing with missing object parts. [9] uses a GentleBoost based classification approach to find candidate points for heart wall landmarks.

1.1 Sparse MRF Appearance Models

Sparse MRF Appearance Models (SAM) are the most closely related approach to the present work. They match shape and appearance models to query images by solving a Markov Random Field. SAMs are based on interest points and an elastic geometric model of their spatial configuration derived from training images and corresponding landmarks (selected interest points). These selection stems either from manual annotations or from a weakly-supervised learning scheme [8]. The appearance of the anatomical structure is encoded through local descriptors around the interest points and along the connecting edges of a Delaunay triangulation. The model thus encompasses information about the mesh topology, mean and standard deviation of edge lengths, circular statistics of edge directions relative to interest point orientation and the local point and edge descriptors.

To localize the structure in a target image, interest points and descriptors are computed. A Markov Random field is set up with as many nodes as there are model landmarks and with the target image interest point IDs as labels. Node probabilities are derived from point descriptor similarities and edge probabilities from the model's edge features. Solving the MRF yields the most probable match of the model onto the target image.

Despite good results several issues remain. The requirement for a priori chosen interest point detectors presents a delicate design choice. Depending on the structure, different interest point detectors may be needed for the different subparts, as shown in [7]. This greatly increases the number of labels for the MRF inhibiting fast inference and increases memory requirements. One of the prime obstacle for the application of this approach to 3D data is the typically overwhelming number of detected interest points, rendering the straight forward application of the approach unfeasible. Consequently, SAMs have not yet been applied to 3D data sets.

Contribution. The contribution of this paper is a method that learns 3D landmark appearance from training data and computes interest points for each model node of a 3D deformable model. Edge length and orientation probabilities are modeled in a novel, combined representation. The multitude of resulting potential candidate configurations is disambiguated by discrete optimization, yielding a localization of complex and repetitive anatomical structures in a target volume.

The paper is structured as follows: In Sec. 2 we outline how to derive application specific interest points. Sec. 3 details how these target point candidates and the geometric model are combined into a graphical model to perform the matching. In Sec. 4 we present the experimental evaluation of our approach, followed by a conclusion and an outlook in Sec. 5.

2 Domain Specific Interest Points

Instead of using one or several fixed interest point detectors we train a Random Forest classifier on Haar wavelet-like descriptors around the model landmarks. Sampling the resulting classification volumes into point sets and clustering them yields target point candidates specific for each landmark. An overview of the method is depicted in Fig. 2. We derive interest points with support from small, local regions which represent an important feature that will be used as unary node costs during discrete optimization. Additionally, due to the clustering bandwidth it avoids the need for non-maxima suppression and reduces the number of candidate interest points.

2.1 Haar-Like Features

For describing the local appearance around the model landmarks we employ a set of 3D features computed using a basis of filters similar to Haar wavelets. These features [16] can be computed using integral volumes [11] in a highly efficient

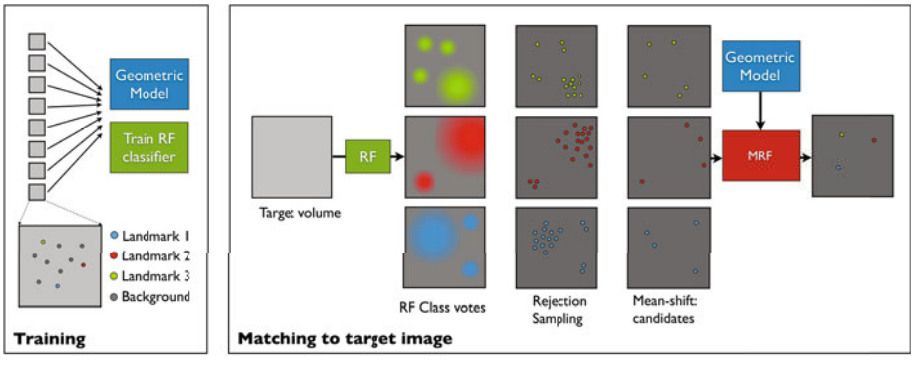


Fig. 2. Outline of the proposed interest point detection and model matching approach

manner. An integral volume $\mathbf{J}$ transforms the information content of the original volume $\mathbf{I}$ such that

$$\mathbf{J}(x,y,z) = \sum_{i=1\ldots x} \sum_{j=1\ldots y} \sum_{k=1\ldots z} \mathbf{I}(i,j,k) \tag{1}$$

This allows to compute the sum s within a cuboid given by the coordinates $(x_1, y_1, z_1, x_2, y_2, z_2)$ by

$$\begin{aligned} s = \mathbf{J}(x_2,y_2,z_2) - \mathbf{J}(x_2,y_1,z_2) - \mathbf{J}(x_1,y_2,z_2) \\ +\mathbf{J}(x_1,y_1,z_1) - \mathbf{J}(x_2,y_2,z_1) + \mathbf{J}(x_2,y_1,z_1) \\ +\mathbf{J}(x_1,y_2,z_1) - \mathbf{J}(x_1,y_1,z_1). \end{aligned} \tag{2}$$

Computing a Haar-like feature then consists of forming 2 such sums with opposing signs. For each of the 3 dimensions, both gradient and ridge features where used, which, together with an average feature derived over the same area as the Haar-features formed the 7 dimensional feature vector for a given wavelet width. 3 different widths wheres used, namely 8, 16 and 32 voxels, yielding a description vector $\mathbf{f}_j$ of dimension 21 for each voxel j in the volume.

2.2 Random Forest Based Appearance Learning and Search

Random Forests [15] are ensemble classifiers. They learn a set of decision trees by randomly sampling from training feature vectors and corresponding labels. During search the decision trees vote for a class label for each query feature vector. Given L landmarks with known positions $\mathbf{x}_l^i$ in all T training volumes ($i = 1, 2, \ldots, T$, $l = 1, 2, \ldots, L$). We assign each training landmark a class label l. In all training volumes we extract local descriptors $\mathbf{f}_j^l$ for all voxels within a 3-voxel radius around landmarks and assign them the corresponding landmark label l. Additionally, we compute descriptors for random background voxels $\mathbf{f}_j^b$. This yields a training set of descriptors, and corresponding labels (L landmark classes, and one background class). A random forest is learnt on the entire set of training examples.

During search on a new target volume, descriptors $\mathbf{f}_j$ are computed for every voxel, and all voxels are classified by the Random Forest. The Random Forests votes are normalized for each class, and yield L volumes $\mathbf{C}^l$ containing the classification probabilities for class l in each voxel. The next step is the generation of landmark candidates from those volumes.

2.3 Mean-Shift Based Interest Point Generation

Mean-shift [3,4] is a method for the density estimation and cluster analysis of a sparse set of points in a, potentially high-dimensional, feature space. Given a d-dimensional dataset $\mathbf{D}$ mean-shift iteratively moves each data point $\mathbf{d}_i$ towards the mean of the data points within a certain distance or bandwidth b (according to a chosen kernel) around $\mathbf{d}_i$. The process is repeated until equilibrium is reached, i. e. when no data point is shifted anymore.

To find the most probable candidates for each model landmark l in the test volume we search for regions within the classification probability volume $\mathbf{C}^l$ with high local support, i. e. regions where their Gaussian weighted integral yields high values. We estimate this by employing mean-shift with a Gaussian kernel with the empirically derived bandwidth $\sigma = 8$ voxels on a rejection sampled version of $\mathbf{C}^l$, i. e. a point set containing the voxels where $\mathbf{C}_l(x, y, z) > r_j$ with r_j being randomly chosen from a uniform distribution between 0 and 1 for each voxel. Each point furthermore carries its probability as weight, leading to means during the mean-shift weighted by $\mathbf{C}_l(x, y, z)$ and the Gaussian kernel.

The result of this process are sets of cluster centers, i. e. interest points, or candidates, $\mathbf{p}_i^l$ for each model landmark l together with estimates of their local support s_i^l consisting of the number of points converged to the cluster center.

3 3D Geometric Model Matching Using Discrete Optimization

In the previous section we have described how to obtain landmark candidates in a search volume. The candidate generation process is based purely on the local appearance learnt from the training cases. In this section, we use the spatial configuration of the landmarks, to constrain the set of landmark candidates, and ultimately find a highly probable landmark assignment in the search volume. The assignment is a tradeoff between local appearance at the landmark position, and the plausibility of the spatial configuration.

We derive an elastic geometric model from the training data, together with confidences about point and edge similarity. The local similarities of this learnt model to the interest points found in the query image are encoded in a Markov Random Field (MRF). The solution of the graphical model yields the match of the model to the query image, i. e. the localization of the anatomical structure in the target volume.

3.1 Modeling Edge Length and Orientation

The landmarks are connected by a set of edges. In contrast to [7] in our experiments we do not employ a Delaunay triangulation but used a manually specified connectivity reflecting the anatomical structure as shown in Fig. 1a, which is almost identical to fully connecting each landmark to its anatomically nearest neighbours. To establish an elastic model from the annotated training landmarks [7] uses mean length and standard deviation of the edges together with circular statistics for relative edge orientations. In [7] both features are used individually to compute confidence values in the range 0 to 1 which are combined to yield similarity confidences between training and target edges.

We propose a more principled approach that combines lengths and orientations and yields proper probabilities through density estimation, as depicted in Fig. 3a **(1-4)**.

For each model edge e from model landmark a to b in the T training volumes two sets P_a, P_b containing the $2T$ endpoints $\mathbf{p}_a^t, \mathbf{p}_b^t$ are known **(1)**. Centering all edges yields the normalized endpoints $\hat{\mathbf{p}}_a^t = \mathbf{p}_a^t - \overline{\langle \mathbf{p}_a^t, \mathbf{p}_b^t \rangle}$ **(2)**.

The mean μ_a and the covariance matrix $\mathbf{\Sigma}_a$ of $\hat{\mathbf{p}}_a^t$, $t = 1, \ldots, T$ now represent the combined length and orientation distribution of edge e: The vector from the origin to μ_a equals the expectation of the orientation $\pm\pi$ and half the length of e **(3)**.

(4) To compute the probability of an edge f between two point candidates $\mathbf{p}_a^f, \mathbf{p}_b^f$ in the target image to be an instance of e we first center $\mathbf{p}_a^f, \mathbf{p}_b^f$: $\hat{\mathbf{p}}_a^f, \hat{\mathbf{p}}_b^f = \mathbf{p}_a^f, \mathbf{p}_b^f - \overline{\langle \mathbf{p}_a^f, \mathbf{p}_b^f \rangle}$. Computing non-normalized Gaussian values d_a, d_b

$$d_a = exp(-\frac{1}{2}(\hat{\mathbf{p}}_a^f - \mu_a)^\top {\mathbf{\Sigma_a}}^{-1}(\hat{\mathbf{p}}_a^f - \mu_a)) \tag{3}$$

$$d_b = exp(-\frac{1}{2}(\hat{\mathbf{p}}_b^f - \mu_b)^\top {\mathbf{\Sigma_b}}^{-1}(\hat{\mathbf{p}}_b^f - \mu_b)) \tag{4}$$

and taking the larger of the two values $d_{max}(f, e) = max(d_a, d_b)$ results in the confidence of edge f being from the distribution of edge model e.

3.2 Formulating the MRF

The objective function for matching a model to an example image is

$$Conf(\mathcal{S}) = \sum_{l=1\ldots L} \mathcal{L}(l, \mathcal{S}(l)) + \sum_{e=1\ldots E} \mathcal{E}(e, \mathcal{S}(e)). \tag{5}$$

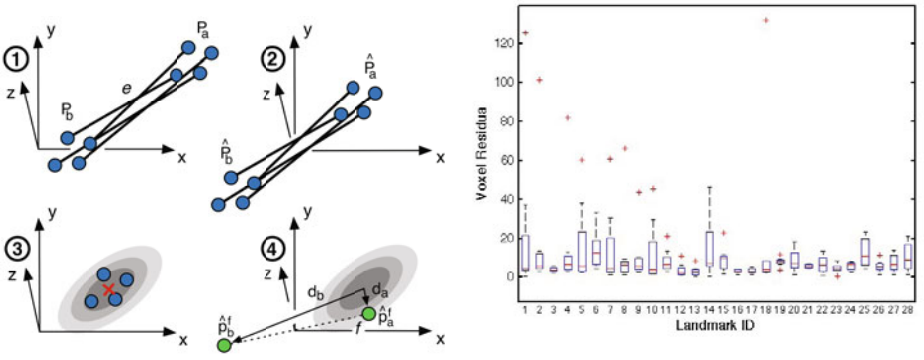
It consists of *unary* terms $\mathcal{L}$ at the nodes of the graphical model describing the L model landmarks to point candidate similarities. *Binary* terms $\mathcal{E}$ capture the similarities of the E model edges to the target edges. Each node has as many labels as that model node has point candidates $\mathbf{p}_i^l$ in the target volume. The confidence for $\mathcal{L}(l, i)$ equals the normalized support $s_i^l / max(s^l)$ and $\mathcal{E}(e, \mathcal{S}(e))$ equals $d_{max}(\mathcal{S}(e), e)$ for the edge between the candidate points $\mathcal{S}(e)$ in relation to model edge e. The MRF's solution, the so called *labeling*

$$\mathcal{S}^* = \underset{\mathcal{S}}{\operatorname{argmax}}\, Conf(\mathcal{S}) \tag{6}$$

assigns each model node l to one point candidate $\mathbf{p}_i^l$ in the target image, matching the model to the target volume.

4 Experiments

We evaluated the proposed approach on 8 Hand CTs with a resolution of 256 × 384 × 330 voxels. 28 landmarks were manually annotated as shown in Fig. 1a. The dataset is challenging due to repetitive nature of the structures. The varying types of structures to be found impede the use of a single interest point detector.



(a) (b)

Fig. 3. (a) **(1-3)** Combined Gaussian model estimation of edge length and orientation from the training instances of edge e. **(4)** Confidence computation that edge f is an instance of model edge e. (b) Residual distances from all leave-one-out runs for each landmark.

In contrast, the Haar-like features together with the Random Forest classifier are well suited to picking up biological structures at different scales, due their detection of salient ridge-like and edge-like features in the volume, like the finger tips and the small joints. Around 50 candidate interest points were detected on average for each landmark. The experiments were run in a leave-one-out cross validation framework using the manual annotations of 7 training volumes to construct the geometric model. The Random Forrest classifier was trained using 200 trees on the 33250 descriptors extracted around the landmarks and from the backgrounds of the training volumes. The MRF was approximately inferred using a simple random walk approach – a detailed comparison of inference methods for this application is subject of ongoing work. After matching, the residual voxel distances between the selected interest points and the corresponding ground truth landmarks were recorded.

Results. To visualize the result quality Fig. 1b shows one of the hands. At each landmark position the radius of a sphere corresponds to the median residual from all leave-one-out runs. Fig. 3b shows the corresponding boxplot.

The mean / median / std residuals measured 10.13 / 5.59 / 16.99 voxels, and 12.87 / 7.01 / 21.57 mm, respectively. They demonstrate the model matching and localization capability of the proposed approach. While the median shows that the vast majority of points are localized with high accuracy (the average finger width being around 32 voxels) 15 outliers considerably deteriorate the mean and standard deviation values. In one instance the volume is cut-off too close to the carpus, leading the solver to choose an alternative point unrelated to the anatomical structure. In 2 instances fingertips crossed over to the adjacent finger. We suspect the small size of the data set to yield a too restrictive geometric model and the MRF solution not representing the global optimum in certain cases.

The runtime of the proposed approach for a single localization, implemented in Matlab except for the C-based RF, amounted to about 5 minutes.

5 Conclusion and Outlook

We present an approach for localizing complex, potentially repetitive anatomical structures in 3D volumes. Based on Random Forests and Haar-like features, the method detects landmark candidate points in a search volume. The anatomical structure is detected by solving a graphical model defined on the landmark candidates. The method does not rely on predefined interest point detectors but derives the most probable candidate locations for each model landmark automatically. This alleviates the prohibitively high number of candidates encountered when using a fixed descriptor together with a simple distance measure. In localization experiments, the method exhibits very low localization errors, but in a few outlier cases single landmark position estimates were attracted to wrong positions.

Future research will focus on increasing the robustness of the approach, especially with regard to matching missing subparts of the objects. While an initial manual indication of the object of interest is deemed necessary, a learning approach should be able to derive which object parts are representative, eliminating the need for detailed annotation. Evaluation on larger data sets and alternative MRF solving strategies will be performed.

References

1. Bergtholdt, M., Kappes, J., Schmidt, S., Schnörr, C.: A study of parts-based object class detection using complete graphs. Int. J. Comput. Vis. 87(1-2), 93–117 (2010)
2. Boykov, Y., Jolly, M.P.: Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In: Proc. ICCV, pp. 105–112 (2001)
3. Cheng, Y.: Mean shift, mode seeking, and clustering. IEEE TPAMI 17(8), 790–799 (1995)
4. Comaniciu, D., Meer, P., Member, S.: Mean shift: a robust approach toward feature space analysis. IEEE TPAMI 24, 603–619 (2002)
5. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE Trans. PAMI 23(6), 681–685 (2001)
6. Criminisi, A., Shotton, J., Bucciarelli, S.: Decision forests with long-range spatial context for organ localization in ct volumes. In: Proc. of MICCAI Workshop on Probabilistic Models for Medical Image Analysis (MICCAI-PMMIA) (2009)
7. Donner, R., Mičušik, B., Langs, G., Bischof, H.: Generalized Sparse MRF Appearance Models (2010)
8. Donner, R., Wildenauer, H., Bischof, H., Langs, G.: Weakly supervised group-wise model learning based on discrete optimization. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 860–868. Springer, Heidelberg (2009)
9. Essafi, S., Langs, G., Paragios, N.: Left ventricle segmentation using diffusion wavelets and boosting. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 919–926. Springer, Heidelberg (2009)

10. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. International Journal on Computer Vision 1, 321–331 (1988)
11. Ke, Y., Sukthankar, R., Hebert, M.: Efficient visual event detection using volumetric features. In: Proc. ICCV (2005)
12. Langs, G., Peloschek, P., Donner, R., Reiter, M., Bischof, H.: Active Feature Models. In: Proc. ICPR, pp. 417–420 (2006)
13. Paragios, N., Deriche, R.: Geodesic Active Contours and Level Sets for the Detection and Tracking of Moving Objects. IEEE PAMI 22(3) (2000)
14. Seifert, S., Barbu, A., Zhou, S.K., Liu, D., Feulner, J., Huber, M., Suehling, M., Cavallaro, A., Comaniciu, D.: Hierarchical parsing and semantic navigation of full body CT data (2009)
15. Statistics, L.B., Breiman, L.: Random forests. In: Machine Learning, pp. 5–32 (2001)
16. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features, pp. 511–518 (2001)
17. Zhan, Y., Zhou, X.S., Peng, Z., Krishnan, A.: Active scheduling of organ detection and segmentation in whole-body medical images. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 313–321. Springer, Heidelberg (2008)
18. Zheng, Y., Georgescu, B., Ling, H., Zhou, S., Scheuering, M., Comaniciu, D.: Constrained marginal space learning for efficient 3d anatomical structure detection in medical images, pp. 194–201 (2009)

Detection of 3D Spinal Geometry Using Iterated Marginal Space Learning

B. Michael Kelm[1], S. Kevin Zhou[2], Michael Suehling[2], Yefeng Zheng[2], Michael Wels[1], and Dorin Comaniciu[2]

[1] Corporate Technology, Siemens AG, Erlangen, Germany
michael.kelm@siemens.com
[2] Siemens Corporate Research, Princeton, USA

Abstract. Determining spinal geometry and in particular the position and orientation of the intervertebral disks is an integral part of nearly every spinal examination with Computed Tomography (CT) and Magnetic Resonance (MR) imaging. It is particularly important for the standardized alignment of the scan geometry with the spine. In this paper, we present a novel method that combines Marginal Space Learning (MSL), a recently introduced concept for efficient discriminative object detection, with a generative anatomical network that incorporates relative pose information for the detection of multiple objects. It is used to simultaneously detect and label the intervertebral disks in a given spinal image volume. While a novel iterative version of MSL is used to quickly generate candidate detections comprising position, orientation, and scale of the disks with high sensitivity, the anatomical network selects the most likely candidates using a learned prior on the individual nine dimensional transformation spaces. Since the proposed approach is learning-based it can be trained for MR or CT alike. Experimental results based on 42 MR volumes show that our system not only achieves superior accuracy but also is the fastest system of its kind in the literature – on average, the spinal disks of a whole spine are detected in 11.5s with 98.6% sensitivity and 0.073 false positive detections per volume. An average position error of 2.4mm and angular error of 3.9° is achieved.

1 Introduction

Examinations of the vertebral column with both Magnetic Resonance (MR) imaging and Computed Tomography (CT) require a standardized alignment of the scan geometry with the spine. While in MR the intervertebral disks can be used to align slice groups and to position saturation bands, in CT the reconstruction planes need to be aligned. In addition to the position and orientation of the disks, physicians are interested in labeling them (e.g. C2/C3, C5/T1, L1/L2, ...). Such a labeling allows to quickly determine the anatomical location without error-prone counting. As manual alignment is both time-consuming and operator-dependent, it is desirable to have a robust, fully automatic, and thus reproducible approach.

B. Menze et al. (Eds.): MICCAI 2010 Workshop MCV, LNCS 6533, pp. 96–105, 2011.

An automatic procedure for extracting the spinal geometry faces various challenges, however. Varying contrasts and image artifacts can compromise the detection of intervertebral disks based on local image features. Thus, a global spine model is required to robustly identify individual disks from their context. Such a model must also cope with missed detections and patients with an unusual number of vertebrae. Finally, the overall approach should run within seconds to allow clinical application.

In this paper we propose a novel approach that combines efficient local object detection based on Marginal Space Learning (MSL) [14] with a global probabilistic model that incorporates pose priors on the nine dimensional parameter spaces that encode position, orientation and scale of the individual disks. The whole approach follows the database-guided detection paradigm [4] and can thus be easily trained for spine detection in CT as well as MR acquired with different sequences.

1.1 Related Work

Recently, the detection and analysis of spinal geometry has regained interest. Boisvert et al. [1] present a model that describes the statistical variations of the spine in terms of sequential rigid transformations of the local vertebra coordinate systems. Using principal component analysis on the Riemannian manifold of rigid transformations they can extract clinically meaningful eigenmodes. Although relying on the same metrics we formulate a probabilistic spine model that is applied for detection rather than statistical analysis.

The detection of intervertebral disks in 3D MR scout scans has recently been addressed by Pekar et al. [8]. They propose a three-step approach using a special-purpose 2D image filter for disk candidate detection, followed by a customized spine tracking method and a final labeling step based on counting. Since their approach is designed to work on MR data only, it might not be easily adapted to CT image volumes.

Schmidt et al. [10] propose a trainable approach based on extremely randomized trees in combination with a complete graphical model. They employ an A*-search based inference algorithm for exact maximum a posteriori (MAP) estimation. The approach only considers the position of the intervertebral disks, while we also determine their orientations and scales. However, their parts-based 3D approach appears most related to ours and their results based on 3D T_1-weighted composed multi-station MR data can best be compared with ours.

Corso et al. [2] argue that a two-level probabilistic model is required to separate pixel-level properties from object-level geometric and contextual properties. They propose a generative graphical model with latent disk variables which they solve by generalized expectation maximization (EM). Although the approach only provides position estimates and has only been evaluated for lumbar disks in 2D T_2-weighted MR data, it could in principle be extended to full 3D estimation. But since EM only finds a local optimum of the expected log likelihood, which can render such an approach very sensitive to initialization, it is not clear how the approach would scale to higher-dimensional estimation including 3D position, orientation, and scale.

2 Methods

Our approach can be subdivided into three major steps (cf. Fig. 1). To constrain the search range for the disks, the spine is roughly located within the given volume first. Second, disk candidates are generated with high sensitivity using a novel iterative extension of the MSL approach [14]. Finally, a global probabilistic spine model is used to select the most likely disk candidates based on their appearance and relative pose and to determine the appropriate label for each disk.

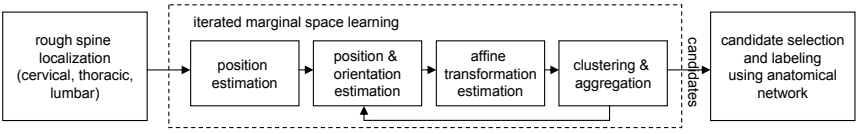


Fig. 1. Overall approach

2.1 Global Probabilistic Spine Model

The typical spatial structure of the spine gives rise to a prior on the relative poses of the spinal disks. This has been modeled by the factor graph [7] depicted in Fig. 2. We have chosen a chain model with potentials considering position, orientation and scale of the spinal disks. Each of the (vector-valued) random variables $\boldsymbol{b}_1$ to $\boldsymbol{b}_N$ represents the pose of a certain spinal disk, thus $\boldsymbol{b}_s$ holds a 3D position $\boldsymbol{p}_s = [x_s, y_s, z_s]^T$, a unit quaternion $\boldsymbol{q}_s$ representing the orientation [6] and an anisotropic scale $\boldsymbol{s}_s = [s_s^x, s_s^y, s_s^z]^T$ for every disk $s \in \{1, \ldots, N\}$. Thus, a distribution over disk poses is defined by the log probability

$$\log \Pr(\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_N \,|\, \boldsymbol{\Theta}, \boldsymbol{I}) = \sum_s V_s(\boldsymbol{b}_s \,|\, \boldsymbol{\theta}_s, \boldsymbol{I}) + \sum_{s \sim t} V_{st}(\boldsymbol{b}_s, \boldsymbol{b}_t \,|\, \boldsymbol{\theta}_{st}) - A \quad (1)$$

where A is the log partition function, $\boldsymbol{I}$ represents the image data and $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_s, \boldsymbol{\theta}_{st}\}$ subsumes all model parameters which are detailed in the following.

The pair potential between two neighboring disk $\boldsymbol{b}_s$ and $\boldsymbol{b}_t$ combines relative position, relative orientation and relative scale terms:

$$V_{st}(\boldsymbol{b}_s, \boldsymbol{b}_t \,|\, \boldsymbol{\theta}_{st}) = V_{pos,st} + V_{rot,st} + V_{sca,st} \quad (2)$$

Each of the terms is defined as a Gaussian pair potential, i.e.,

$$V_{pos,st}(\boldsymbol{b}_s, \boldsymbol{b}_t) = -\frac{1}{2}\, \boldsymbol{d}_{pos}^T(\boldsymbol{b}_s, \boldsymbol{b}_t)\, \boldsymbol{\Sigma}_{pos,st}^{-1}\, \boldsymbol{d}_{pos}(\boldsymbol{b}_s, \boldsymbol{b}_t) \quad (3)$$

$$V_{rot,st}(\boldsymbol{b}_s, \boldsymbol{b}_t) = -\frac{\alpha(\boldsymbol{q}_t \boldsymbol{q}_s^{-1} \boldsymbol{\mu}_{rot,st}^{-1})^2}{2\sigma_{rot,st}^2} \quad (4)$$

$$V_{sca,st}(\boldsymbol{b}_s, \boldsymbol{b}_t) = -\frac{1}{2}\, \boldsymbol{d}_{sca}^T(\boldsymbol{b}_s, \boldsymbol{b}_t)\, \boldsymbol{\Sigma}_{sca,st}^{-1}\, \boldsymbol{d}_{sca}(\boldsymbol{b}_s, \boldsymbol{b}_t) \quad (5)$$

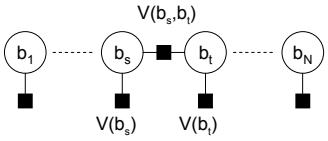


Fig. 2. Factor graph modeling the relation between the spinal disks

with the rotation angle $\alpha(\boldsymbol{q}) = \alpha([q_0\, q_1\, q_2\, q_3]) = 2\arccos(q_0)$ and with $\boldsymbol{d}_{pos}(\boldsymbol{b}_s, \boldsymbol{b}_t) = \mathbf{R}_s^{-1}(\boldsymbol{p}_t - \boldsymbol{p}_s) - \boldsymbol{\mu}_{pos,st}$ and $\boldsymbol{d}_{sca}(\boldsymbol{b}_s, \boldsymbol{b}_t) = \boldsymbol{s}_t - \boldsymbol{s}_s - \boldsymbol{\mu}_{sca,st}$ where $\mathbf{R}_s$ is the rotation matrix associated with the quaternion $\boldsymbol{q}_s$. In summary, the pair potential parameters $\boldsymbol{\theta}_{st}$ are the mean parameters $\boldsymbol{\mu}_{pos,st}$, $\boldsymbol{\mu}_{rot,st}$, $\boldsymbol{\mu}_{sca,st}$ and the (co-)variance parameters $\boldsymbol{\Sigma}_{pos,st}$, $\sigma_{rot,st}$, $\boldsymbol{\Sigma}_{sca,st}$. To keep the number of estimated parameters small, both $\boldsymbol{\Sigma}_{pos,st}$ and $\boldsymbol{\Sigma}_{sca,st}$ are constrained to diagonal matrices.

Both, the position and the scale potentials are defined based on Euclidean distance. The required mean parameters $\boldsymbol{\mu}_{pos,st}$ and $\boldsymbol{\mu}_{sca,st}$ and the covariance matrices $\boldsymbol{\Sigma}_{pos,st}$ and $\boldsymbol{\Sigma}_{sca,st}$ are determined from the training data. The rotation potential in Eqn. (4) uses the intrinsic metric $\alpha(\boldsymbol{q})$ of the corresponding manifold $\mathcal{SO}^3$. Consequently, the mean rotation is determined as the Fréchet mean [9]. Collecting all instances of a certain disk pair $(\boldsymbol{b}_s, \boldsymbol{b}_t)$ into the training sample $\mathcal{P}_{st}$, the Fréchet mean for the corresponding rotation potential is determined as

$$\boldsymbol{\mu}_{rot,st} = \underset{|\boldsymbol{q}|=1}{\operatorname{argmin}} \sum_{(\boldsymbol{b}_s,\boldsymbol{b}_t)\in\mathcal{P}_{st}} \alpha(\boldsymbol{q}_t \boldsymbol{q}_s^{-1} \boldsymbol{q}^{-1})^2. \tag{6}$$

It can be efficiently computed using the eigen-decomposition proposed in reference [6]. The Gaussian variance is estimated with

$$\sigma_{rot,st}^2 = \frac{1}{|\mathcal{P}_{st}| - 1} \sum_{(\boldsymbol{b}_s,\boldsymbol{b}_t)\in\mathcal{P}_{st}} \alpha(\boldsymbol{q}_t \boldsymbol{q}_s^{-1} \boldsymbol{\mu}_{rot,st}^{-1})^2. \tag{7}$$

Finally, the single site potentials, which are determined by iterated marginal space learning as described in the following section, encode image-based likelihood, i.e.,

$$V_s(\boldsymbol{b}_s \,|\, \boldsymbol{\theta}_s, \boldsymbol{I}) = log(\Pr(\boldsymbol{b}_s \,|\, \boldsymbol{\theta}_s, \boldsymbol{I})). \tag{8}$$

Since the defined potentials are invariant under global rigid transformations (translation and rotation), the resulting distribution is insensitive towards different poses of the spine. Furthermore, models capturing only parts of the complete spine can be easily constructed by just omitting the superfluous disk variables. Since all potential parameters are determined independently (i.e., the likelihood decouples), no retraining is required and a probabilistic model appropriate for the current acquisition protocol, e.g., a lumbar spine protocol, can be assembled at runtime.

2.2 Iterated Marginal Space Learning

In principle, the defined potentials can be evaluated for every possible position, orientation and scale. However, performing an exhaustive search on the uniformly discretized nine dimensional parameter space (3 position, 3 orientation and 3 scale parameters) would require evaluating a huge number of single site as well as pair potentials. Such a direct approach would be computationally very expensive.

Hence, we adopt the MSL paradigm [14], a novel concept that has recently proven successful in numerous applications [3,5,13]. Instead of searching the whole nine dimensional parameter space, the MSL paradigm proposes a three-step approach. First candidate positions for the sought object are collected by using a probabilistic machine learning classifier to check every voxel location within a defined range. In the second step, a number of 3D orientation hypotheses that have been derived from the training set are evaluated by a second classifier using the the most likely object positions from the first step. Similarly, the last step estimates three scale parameters based on the candidates from the second step using a third classifier.

MSL has been designed to detect a single, specific object such as, for example, a particular organ or landmark. If multiple objects of the same type are to be detected, as in our case, the described MSL approach may end up with detections for the most salient disks only, i.e., many disks would be missed. Although the sensitivity could be improved by drastically increasing the number of considered candidates in each step, this is not practicable since MSL would then loose its computational efficiency.

We therefore propose a novel extension to MSL, iterative MSL (iMSL), to cope with multiple objects of the same type (cf. Fig. 3). It is designed to achieve a higher sensitivity than usual MSL at moderate computational costs. First, the position detector is evaluated in each voxel of the given image volume region. The N_0 most likely candidates are collected in the set of initial position candidates $\mathcal{P}_0$. Then, the best N_{pos} ($N_{pos} < N_0$) candidates from $\mathcal{P}_0$ are evaluated using the orientation detector whose top candidates are evaluated using the scale detector. The resulting set $\mathcal{D}_{sca}$ contains disk candidate detections with all estimated parameters. Using pairwise average-linkage clustering with Euclidean distance, clusters of candidate disks are obtained. The most likely N_A box candidates of each resulting cluster are averaged and added to the set of detected disk candidates $\mathcal{D}$. After removing all position candidates from $\mathcal{P}_0$ that are closer than a specified radius R to any of the detections in $\mathcal{D}$, orientation and scale detection are repeated on the remaining position candidates until no position candidates are left or no new disk candidates are detected.

Like Zheng et al. [14] we employ the probabilistic boosting tree (PBT) classifier using Haar-like features for the position detector and steerable features for the orientation and scale detectors.

The probabilistic spine model described in the previous section is discretized using the disk candidates detected with iMSL. Each random variable $\boldsymbol{b}_s$ is transformed into a discrete random variable where each state represents one of the detected disk candidates. In order to allow for missed detections, an extra

Input: R, N_0, N_{pos}, N_{ort}, N_{sca}
Output: Set $\mathcal{D}$ of detected disk candidates
$\mathcal{D} := \{\}$;
$\mathcal{P}_0 :=$ the N_0 most likely candidates according to the position detector;
repeat
  $\mathcal{P}_0 := \{p \in \mathcal{P}_0 : d(p,q) > R \; \forall q \in \mathcal{D}\}$;
  $\mathcal{D}_{pos} :=$ the N_{pos} most likely candidates from $\mathcal{P}_0$;
  $\mathcal{D}_{ort} :=$ the N_{ort} most likely candidates from $\mathcal{D}_{pos}$ according to the orientation detector;
  $\mathcal{D}_{sca} :=$ the N_{sca} most likely candidates from $\mathcal{D}_{ort}$ according to the scale detector;
  Perform hierarchical agglomerative clustering on $\mathcal{D}_{sca} \cup \mathcal{D}$;
  foreach *cluster* $\mathcal{C}$ **do**
    if $|\mathcal{C}| \geq N_A$ **then**
      Aggregate the top N_A candidates and add the resulting box to $\mathcal{D}$;
    end
  end
until $|\mathcal{P}_0| = 0$ *or* $|\mathcal{D}|$ *remains constant*;

Fig. 3. Pseudo-code for Iterated Marginal Space Learning (iMSL)

"missing" state is introduced. Note, that iMSL detects disk candidates with high sensitivity which usually results in more disk candidates than actual disks. The MAP estimate, i.e. the maximum of Eqn. (1)), provides the optimum assignment of a disk candidate to one of the disk variables according to the probabilistic spine model. Thus, only those disk candidates that form a valid spine are selected and are implicitly assigned a suitable label.

The MAP is efficiently computed by belief propagation where, due to the tree structure of the factor graph (cf. Fig. 2), a single forward-backward pass yields the exact solution [7]. An additional speed-up is obtained by constraining the search for disk candidates to the area of the spine. For this purpose, bounding boxes around the lumbar, thoracic and cervical regions of the spine are detected first using the usual MSL approach as described in reference [14].

3 Experimental Results

3.1 Data

Experiments have been conducted based on 3D T_1-weighted MR volumes (FL3D-VIBE sequence) from 42 volunteers. About one half of the volumes has been acquired on two 1.5T scanner models (MAGNETOM Avanto and MAGNETOM Espree, Siemens AG, Erlangen) with TR = 5/4ms, TE = 2ms and a flip angle of 10°. The other half has been obtained from two 3T scanner models (MAGNETOM Trio, MAGNETOM Verio, Siemens AG, Erlangen) with TR = 4/3ms,

TE = 1ms and again a flip angle of 10°. Each of the volumes was recorded in a two station scan and subsequently combined to a volume covering the whole spine (approximately 860mm × 350mm × 190mm) with an isotropic resolution of 2.1mm. Susceptibility artifacts and intensity variations due to magnetic field inhomogeneities were present in the data. No bias field correction was performed.

3.2 Results

To obtain ground truth, each intervertebral disk has been annotated with four defined landmarks. From these, ground truth boxes have been derived for the intervertebral disks as well as the lumbar, thoracic and cervical spine regions. For disk detection, *i*MSL was employed with a cluster radius of $R = 6$mm, $N_0 = 3000$ initial position candidates and 500 detection candidates for the remaining detection estimation steps ($N_{pos} = 500$, $N_{ort} = 500$, $N_{sca} = 500$).

All evaluation results have been obtained using 10-fold cross validation, ensuring that training and testing data never stem from the same patient. Every ground truth annotation for which no disk within a distance of 10mm was detected, was counted as a missed detection. Overall, intervertebral disks have been detected with a sensitivity of 98.64% and only 0.0731 false positives per volume, yielding a positive predictive value of 99.68%. The overall processing time on a 2.2GHz dual core laptop computer was between 9.9s and 13.0s and 11.5s on average where most of the time was spent on disk candidate detection.

The accuracy of the detected intervertebral disks has been evaluated by the position distance and the angle between the disk plane normals of the detected intervertebral disks and the ground truth annotation (cf. Table 1). On average, a position error of 2.42mm (about 1 voxel) and an angular error of 3.85° was obtained.

Table 1. Disk detection results using 10-fold cross validation based on 42 T_1-weighted MR volumes. **Left**: position error [mm]. **Right**: angular error between normals [degree].

	cervical	thoracic	lumbar	overall	cervical	thoracic	lumbar	overall
mean	2.09	2.41	2.86	2.42	4.86	3.38	3.80	3.85
median	1.84	2.18	2.68	2.19	3.89	2.90	3.37	3.17
lower quartile	1.40	1.56	1.88	1.58	2.52	1.85	2.08	1.97
upper quartile	2.63	3.00	3.63	3.05	6.68	4.48	5.03	5.02

Four examples from the MR data set are shown in Fig. 4. The right-most example shows a case where the volunteer has been instructed to lie down twisted in order to simulate a scoliotic spine. Still the proposed approach could locate and label all spinal disks reliably.

Some results on lumbar spine CT are shown in Fig. 5. While the complete probabilistic spine model as well as the *i*MSL detectors have been trained on CT data showing various regions of the spine, our approach allows to assemble an appropriate model for the lumbar spine without retraining.

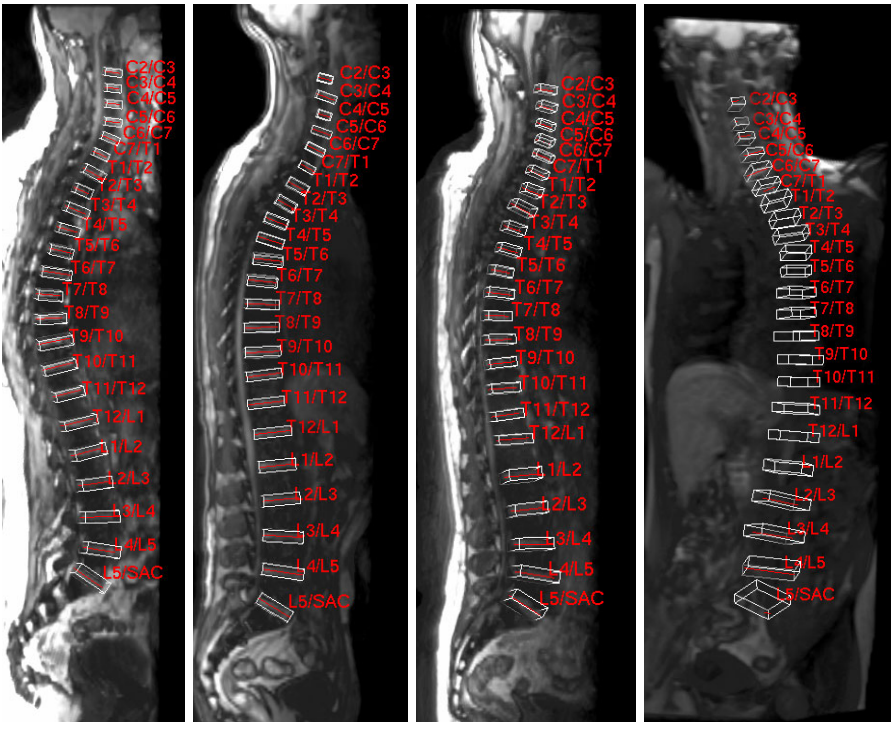


Fig. 4. Four examples from the MR data with detection results. Although the volunteer in the rightmost example lay down in an unusually twisted pose, all intervertebral disks were detected and labeled correctly.

Fig. 5. Detection results for CT scans of the lumbar spine

The results of our proposed method compare favorably with results presented in previous works. While with only 6s processing time the approach by Pekar et al. [8] runs faster than ours, it has lower sensitivity (95.6% before candidate selection) and does not provide orientation estimates.

Compared with the best cross validation results by Schmidt et al. [10], the results obtained with our approach are significantly better. While a competitive but still smaller sensitivity of 97% is reported, they only achieve a position error of 5.1mm. Furthermore, no orientation estimates are provided and the approach takes several minutes to run. Furthermore, in contrast to Schmidt et al. [10], we did not perform any posterior search at the positions of missing disks which could further increase our sensitivity.

While at the current state we did not perform systematic testing on data from patients with pathologies (e.g. scoliosis, stenosis, disk degeneration, herniation, desiccation), we are confident that our approach also works for disease cases. In this as well as other applications we have observed, that the MSL approach is very robust to imaging artifacts and unusual appearances of the sought object. Using *i*MSL, increases sensitivity and helps detect disks with very unusual appearance. Furthermore, since the global spine model is restricted to candidates provided by the disk detector, scoliotic abnormalities can be robustly handled. The volunteer with the twisted pose in Fig. 4 provides evidence towards this. Finally, simple retraining of our system with some abnormal cases added, enables the detectors as well as the prior model to handle them even more reliably.

4 Conclusion and Future Work

In this paper, we have presented a novel approach to the fully automatic detection of 3D spinal geometry and labeling of the intervertebral disks. The approach uses an iterative extension of MSL for disk candidate detection along with an anatomical network that incorporates spatial context in form of a prior on the nine dimensional disk poses. Since the entire approach is learning-based, it can be trained for CT and MR alike.

Using 42 MR image volumes, superior sensitivity and accuracy was obtained than in previous works. With an overall processing time of only 11.5s, the approach is also comparably fast and can be used as routine procedure for the automatic planning of scan geometries. Results on CT data show that the proposed approach can be adapted to different modalities. For this purpose, the graphical model can be adjusted to handle partial spine recordings that are commonly acquired with CT.

Apart from automatic scan alignment, the proposed system for detecting and labeling the intervertebral disks could be part of a computer-aided diagnosis system for analyzing pathologies of the intervertebral disks or the vertebrae. The detected bounding boxes could, for example, be used for initializing a detailed vertebra segmentation algorithm with subsequent analysis. Furthermore, the proposed system could support semantic body parsing and semantic annotation to automatically generate semantic location descriptions as frequently used by physicians for reporting [12,11]. Both applications will be considered in future work.

References

1. Boisvert, J., Cheriet, F., Pennec, X., Labelle, H., Ayache, N.: Geometric variability of the scoliotic spine using statistics on articulated shape models. IEEE Trans. Med. Imag. 27(4), 557–568 (2008)
2. Corso, J.J., Alomari, R.S., Chaudhary, V.: Lumbar disc localization and labeling with a probabilistic model on both pixel and object features. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241,pp. 202–210. Springer, Heidelberg (2008)
3. Feng, S., Zhou, S., Good, S., Comaniciu, D.: Automatic fetal face detection from ultrasound volumes via learning 3D and 2D information. In: Proc. CVPR, pp. 2488–2495 (2009)
4. Georgescu, B., Zhou, X.S., Comaniciu, D., Gupta, A.: Database-guided segmentation of anatomical structures with complex appearance. In: Proc. CVPR, pp. 429–436 (2005)
5. Ionasec, R.I., Voigt, I., Georgescu, B., Wang, Y., Houle, H., Hornegger, J., Navab, N., Comaniciu, D.: Personalized modeling and assessment of the aortic-mitral coupling from 4D TEE and CT. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 767–775. Springer, Heidelberg (2009)
6. Karney, C.F.: Quaternions in molecular modeling. J. Molec. Graph. Modelling 25, 595–604 (2007)
7. Kschischang, F.R., Frey, B.J., Loeliger, H.A.: Factor graphs and the sum-product algorithm. IEEE Trans. Inf. Theory 47(2), 498–519 (2001)
8. Pekar, V., Bystrov, D., Heese, H.S., Dries, S.P.M., Schmidt, S., Grewer, R., den Harder, C.J., Bergmans, R.C., Simonetti, A.W., van Muiswinkel, A.M.: Automated planning of scan geometries in spine MRI scans. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 601–608. Springer, Heidelberg (2007)
9. Pennec, X.: Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. J. Math. Imag. Vis. 25(1), 127–154 (2006)
10. Schmidt, S., Kappes, J., Bergtholdt, M., Pekar, V., Dries, S., Bystrov, D., Schnörr, C.: Spine detection and labeling using a parts-based graphical model. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 122–133. Springer, Heidelberg (2007)
11. Seifert, S., Barbu, A., Zhou, S.K., Liu, D., Feulner, J., Huber, M., Sühling, M., Cavallaro, A., Comaniciu, D.: Hierarchical parsing and semantic navigation of full body ct data. In: Proc. SPIE Medical Imaging, pp. 725–732 (2009)
12. Seifert, S., Kelm, M., Möller, M., Mukherjee, S., Cavallaro, A., Huber, M., Comaniciu, D.: Semantic annotation of medical images. In: Proc. SPIE Medical Imaging. o.A. (2010)
13. Wels, M., Zheng, Y., Carneiro, G., Huber, M., Hornegger, J., Comaniciu, D.: Fast and robust 3-D MRI brain structure segmentation. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 575–583. Springer, Heidelberg (2009)
14. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. IEEE Trans. Med. Imag. 27(11), 1668–1681 (2008)

Regression Forests for Efficient Anatomy Detection and Localization in CT Studies

Antonio Criminisi, Jamie Shotton, Duncan Robertson, and Ender Konukoglu

Microsoft Research Ltd, CB3 0FB, Cambridge, UK

Abstract. This paper proposes multi-class random regression forests as an algorithm for the efficient, automatic detection and localization of anatomical structures within three-dimensional CT scans.

Regression forests are similar to the more popular classification forests, but trained to predict *continuous* outputs. We introduce a new, continuous parametrization of the anatomy localization task which is effectively addressed by regression forests. This is shown to be a more natural approach than classification.

A single pass of our probabilistic algorithm enables the direct mapping from voxels to organ location and size; with training focusing on maximizing the confidence of output predictions. As a by-product, our method produces *salient anatomical landmarks*; *i.e.* automatically selected "anchor" regions which help localize organs of interest with high confidence. Quantitative validation is performed on a database of 100 highly variable CT scans. Localization errors are shown to be lower (and more stable) than those from global affine registration approaches. The regressor's parallelism and the simplicity of its context-rich visual features yield typical runtimes of only $1s$. Applications include semantic visual navigation, image tagging for retrieval, and initializing organ-specific processing.

1 Introduction

This paper introduces the use of regression forests in the medical imaging domain and proposes a new, parallel algorithm for the efficient detection and localization of anatomical structures ('organs') in computed tomography (CT) studies.

The main contribution is a new parametrization of the anatomy localization task as a multi-variate, continuous parameter estimation problem. This is addressed effectively via tree-based, non-linear regression. Unlike the popular *classification* forests (often referred to simply as "random forests"), regression forests [1] have not yet been used in medical image analysis. Our approach is fully probabilistic and, unlike previous techniques (*e.g.* [2,3]) maximizes the confidence of output predictions. The focus of this paper is both on accuracy of prediction and speed of execution, as we wish to achieve anatomy localization in seconds. Automatic anatomy localization is useful for efficient visual navigation, initializing further organ-specific processing (*e.g.* detecting liver tumors), and semantic tagging of patient scans to aid their sorting and retrieval.

B. Menze et al. (Eds.): MICCAI 2010 Workshop MCV, LNCS 6533, pp. 106–117, 2011.

Regression-based approaches. Regression algorithms [4] estimate functions which map input variables to *continuous* outputs[1]. The regression paradigm fits the anatomy localization task well. In fact, its goal is to learn the non-linear mapping from voxels *directly* to organ position and size. [5] presents a thorough overview of regression techniques and demonstrates the superiority of boosted regression [6] with respect to *e.g.* kernel regression [7]. In contrast to the boosted regression approach in [2] maximizing confidence of output prediction is integral to our approach. A comparison between boosting, forests and cascades is found in [8]. To our knowledge only two papers have used regression forests [9,10]; neither with application to medical image analysis, nor to multi-class problems. For instance, [10] addresses the problem of detecting pedestrians v background.

Classification-based approaches. In [11] organ detection is achieved via a confidence maximizing sequential scheduling of multiple, organ-specific *classifiers*. Our single, tree-based regressor allows us to deal naturally with multiple anatomical structures simultaneously. As shown in the machine learning literature [12] this encourages feature sharing and, in turn better generalization. In [13] a sequence of PBT classifiers (first for salient slices, then for landmarks) are used. In contrast, our single regressor maps directly from voxels to organ poses. Latent, salient landmark regions are extracted as a by-product of our procedure. In [14] the authors achieve localization of organ *centres* but fail to estimate the organ extent (similarly for [10]). Here we present a more direct, continuous model which estimates the position of the walls of the bounding box containing each organ; thus achieving simultaneous organ localization and extent estimation.

Registration-based approaches. Although atlas-based methods have enjoyed much popularity [3,15,16] their conceptual simplicity is in contrast to the need for robust, cross-patient registration. Robustness is improved by multi-atlas techniques [17], at the price of slower algorithms involving multiple registrations. Our algorithm incorporates atlas information within a compact tree-based model. As shown in the result section, such model is more efficient than keeping around multiple atlases and achieves anatomy localization in only a few seconds. Comparisons with affine registration methods (somewhat similar to ours in computational cost) show that our algorithm produces lower and more stable errors.

1.1 Background on Regression Trees

Regression trees [18] are an efficient way of mapping a complex input space to continuous output parameters. Highly non-linear mappings are handled by splitting the original problem into a set of smaller problems which can be addressed with simple predictors. Figure 1 shows an illustrative 1D example where the goal is to learn an analytical function to predict the real-valued output y (*e.g.* house prices) given the input x (*e.g.* air pollution). Learning is supervised as we are given a set of training pairs (x, y). Each node in the tree is designed to split the data so as to form clusters where accurate prediction can be performed with

[1] As opposed to *classification* where the predicted variables are discrete.

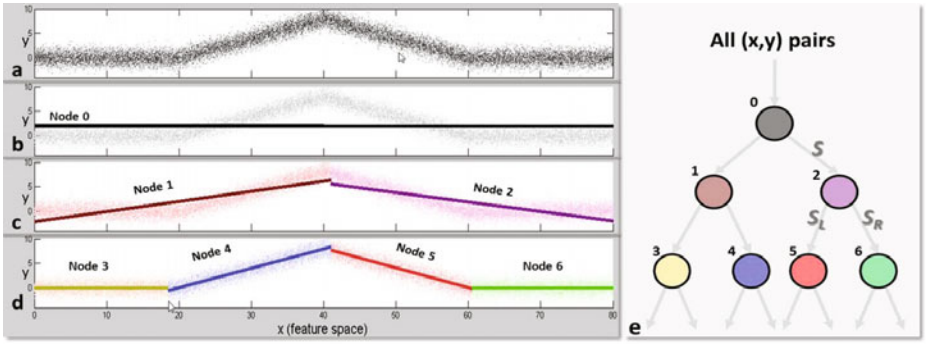


Fig. 1. Regression tree: an explanatory 1D example. **(a)** Input data points. **(b)** A single linear function fits the data badly. **(c,d)** Using more tree levels yields more accurate fit of the regressed model. Complex non-linear mappings are modelled via a hierarchical combination of many, simple linear regressors. **(e)** The regression tree.

simpler models (*e.g.* linear in this example). More formally, each node performs the test $\xi > f(x) > \tau$, with ξ, τ scalars. Based on the result each data point is sent to the left or right child.

During training, each node test (*e.g.* its parameters ξ, τ) is optimized so as to obtain the best split; *i.e.* the split that produces the maximum reduction in geometric error. The error reduction r is defined here as: $r = e(\mathcal{S}) - \sum_{\mathtt{i} \in \mathtt{L},\mathtt{R}} \omega_i e(\mathcal{S}_i)$ where $\mathcal{S}$ indicates the set of points reaching a node, and `L` and `R` denote the left and right children (for binary trees). $\omega_i = |\mathcal{S}_i|/|\mathcal{S}|$ is the ratio of the number of points reaching the i^{th} child. For a set $\mathcal{S}$ of points the error of geometric fit is: $e(\mathcal{S}) = \sum_{j \in \mathcal{S}} [y_j - y(x_j; \boldsymbol{\eta}_S)]^2$, with $\boldsymbol{\eta}_S$ the two line parameters computed from all points in $\mathcal{S}$ (*e.g.* via least squares or RANSAC). Each leaf stores the continuous parameters $\boldsymbol{\eta}_S$ characterizing each linear regressor. More tree levels yield smaller clusters and smaller fit errors, but at the risk of overfitting.

2 Multivariate Regression Forests for Organ Localization

This section presents our mathematical parametrization and the details of our multi-organ regression forest with application to anatomy localization.

Mathematical notation. Vectors are represented in boldface (*e.g.* $\mathbf{v}$), matrices as teletype capitals (*e.g.* $\mathtt{A}$) and sets in calligraphic style (*e.g.* $\mathcal{S}$). The position of a voxel in a CT volume is denoted $\mathbf{v} = (v_x, v_y, v_z)$.

The labelled database. The anatomical structures we wish to recognize are $\mathcal{C} =\{$ `heart, liver, spleen, left lung, right lung, l. kidney, r. kidney, gall bladder, l. pelvis, r. pelvis`$\}$. We are given a database of 100 scans which have been manually annotated with 3D bounding boxes tightly drawn around the structures of interest (see fig. 3a). The bounding box for the organ

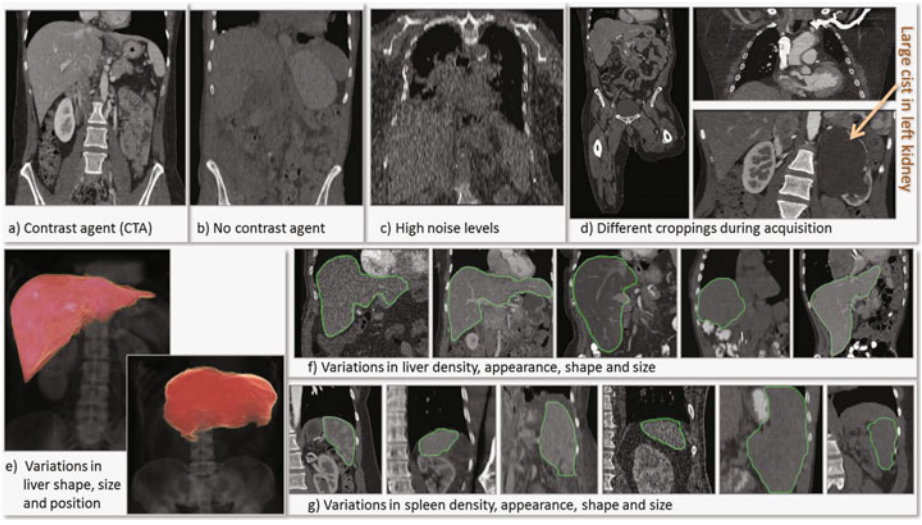


Fig. 2. Variability in our labelled database. (a,b,c) Variability in appearance due to presence of contrast agent, or noise. **(d)** Difference in image geometry due to acquisition parameters and possible anomalies. **(e)** Volumetric renderings of liver and spine to illustrate large changes in their relative position and in the liver shape. **(f,g)** Mid-coronal views of liver and spleen across different scans in our database to illustrate their variability. All views are metrically and photometrically calibrated.

$c \in \mathcal{C}$ is parametrized as a 6-vector $\mathbf{b}_c = (b_c^{\mathtt{L}}, b_c^{\mathtt{R}}, b_c^{\mathtt{A}}, b_c^{\mathtt{P}}, b_c^{\mathtt{H}}, b_c^{\mathtt{F}})$ where each component represents the position (in mm) of the corresponding axis-aligned wall[2]. The database comprises patients with different conditions and large differences in body size, pose, image cropping, resolution, scanner type, and possible use of contrast agents (fig. 2). Voxel sizes are $\sim 0.5 - 1.0mm$ along x and y, and $\sim 1.0 - 5.0mm$ along z. The images have not been pre-registered or normalized in any way. The goal is to localize anatomies of interest accurately and automatically, despite such large variability. Next we describe how this is achieved.

2.1 Problem Parametrization and Regression Forest Learning

Key to our algorithm is the fact that *all* voxels in a test CT volume contribute with varying confidence to estimating the position of the six walls of *all* structures' bounding boxes (see fig. 3b,c). Intuitively, some distinct voxel clusters (*e.g.* ribs or vertebrae) may predict the position of an organ (*e.g.* the heart) with high confidence. Thus, during testing those clusters will be used as reference (landmarks) for the localization of those anatomical structures. Our aim is to learn to cluster voxels together based on their appearance, their spatial context and, above all, their confidence in predicting position and size of all

[2] Superscripts follow standard radiological orientation convention: $\mathtt{L}$ = left, $\mathtt{R}$ = right, $\mathtt{A}$ = anterior, $\mathtt{P}$ = posterior, $\mathtt{H}$ = head, $\mathtt{F}$ = foot.

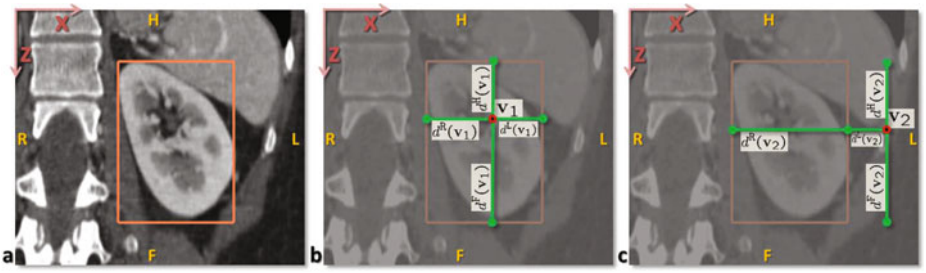


Fig. 3. Problem parametrization. (a) A coronal view of a left kidney and the associated ground-truth bounding box (in orange). **(b,c)** *Every* voxel $\mathbf{v}_i$ in the volume votes for the position of the six walls of each organ's 3D bounding box via 6 relative, offset displacements $d^{\mathrm{k}}(\mathbf{v}_i)$ in the three canonical directions x, y and z.

anatomical structures. We tackle this simultaneous feature selection and parameter regression task with a multi-class random regression forest (fig. 4); *i.e.* an ensemble of regression trees trained to predict location and size of all desired anatomical structures simultaneously.

Note that in the illustrative example in section 1.1 the goal was to estimate a two-dimensional continuous vector representing a line. In contrast, here the desired output is one six-dimensional vector $\mathbf{b}_c$ per organ, for a total of $6|\mathcal{C}|$ continuous parameters. Also note that this is very different from the task of assigning a categorical label to each voxel (*i.e.* the classification approach in [14]). Here we wish to produce confident predictions of a small number of continuous localization parameters. The *latent* voxel clusters are discovered automatically without supervised cluster labels.

Forest training. The training process constructs each regression tree and decides at each node how to best split the incoming voxels. We are given a subset of all labelled CT volumes (the training set), and the associated ground-truth organ bounding boxes (fig. 3a). The size of the forest T is fixed and all trees are trained in parallel. Each voxel is pushed through each of the trees starting at the root. Each split node applies the following binary test $\xi_j > f(\mathbf{v};\boldsymbol{\theta}_j) > \tau_j$ and based on the result sends the voxel to the left or right child node. $f(.)$ denotes the feature response computed for the voxel $\mathbf{v}$. The parameters $\boldsymbol{\theta}_j$ represent the visual feature which applies to the j^{th} node. Our visual features are similar to those in [10,14,19], *i.e.* mean intensities over displaced, asymmetric cuboidal regions. These features are efficient and capture spatial context. The feature response is $f(\mathbf{v};\boldsymbol{\theta}_j) = |F_1|^{-1} \sum_{\mathbf{q}\in F_1} I(\mathbf{q}) - |F_2|^{-1} \sum_{\mathbf{q}\in F_2} I(\mathbf{q})$; with F_i indicating 3D box regions and I the intensity. F_2 can be the empty set for unary features. Randomness is injected by making available at each node only a random sample of all features. This technique has been shown to increase the generalization of tree-based predictors [1]. Next we discuss how to select the node test.

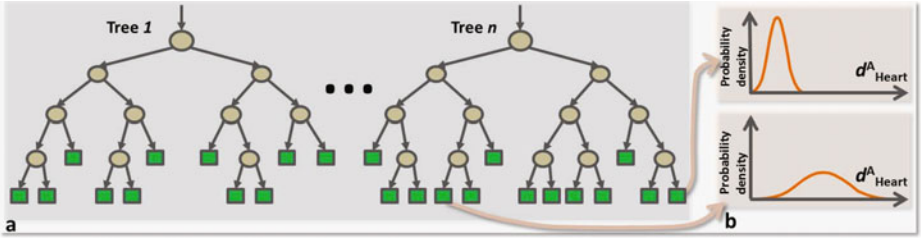


Fig. 4. A regression forest is an ensemble of different regression trees. Each leaf contains a distribution for the continuous output variable/s. Leaves have associated different degrees of confidence (illustrated by the "peakiness" of distributions).

Node optimization. Each voxel $\mathbf{v}$ in each training volume is associated with an offset $\mathbf{d}_c(\mathbf{v})$ with respect to the bounding box $\mathbf{b}_c$ for each class $c \in \mathcal{C}$ (see fig. 3b,c). Such offset is denoted: $\mathbf{d}_c(\mathbf{v}) = (d_c^{\mathtt{L}}, d_c^{\mathtt{R}}, d_c^{\mathtt{A}}, d_c^{\mathtt{P}}, d_c^{\mathtt{H}}, d_c^{\mathtt{F}}) \in \mathbb{R}^6$, with $\mathbf{b}_c(\mathbf{v}) = \hat{\mathbf{v}} - \mathbf{d}_c(\mathbf{v})$ and $\hat{\mathbf{v}} = (v_x, v_x, v_y, v_y, v_z, v_z)$. As with classification, node optimization is driven by maximizing an information gain measure, defined as: $IG = H(\mathcal{S}) - \sum_{i=\{\mathtt{L},\mathtt{R}\}} \omega_i H(\mathcal{S}_i)$ where H denotes entropy, $\mathcal{S}$ is the set of training points reaching the node and $\mathtt{L}, \mathtt{R}$ denote the left and right children. In classification the entropy is defined over distributions of discrete class labels. In regression instead we measure the purity of the probability density of the real-valued predictions. For a single class c we model the distribution of the vector $\mathbf{d}_c$ at each node as a multivariate Gaussian; *i.e.* $p(\mathbf{d}_c) = \mathcal{N}(\mathbf{d}_c; \overline{\mathbf{d}_c}, \Lambda_c)$, with the matrix Λ_c encoding the covariance of $\mathbf{d}_c$ for all points in $\mathcal{S}$. The differential entropy of a multivariate Gaussian can be shown to be $H(\mathcal{S}) = \frac{n}{2}(1 + \log(2\pi)) + \frac{1}{2}\log|\Lambda_c(\mathcal{S})|$ with n the number of dimensions ($n = 6$ in our case). Algebraic manipulation yields the following regression information gain: $IG = \log|\Lambda_c(\mathcal{S})| - \sum_{i=\{\mathtt{L},\mathtt{R}\}} \omega_i \log|\Lambda_c(\mathcal{S}_i)|$. In order to handle simultaneously all $|\mathcal{C}| = 10$ anatomical structures the information gain is adapted to: $IG = \sum_{c\in\mathcal{C}} \left(\log|\Lambda_c(\mathcal{S})| - \sum_{i=\{\mathtt{L},\mathtt{R}\}} \omega_i \log|\Lambda_c(\mathcal{S}_i)|\right)$ which is readily rewritten as

$$IG = \log|\Gamma(\mathcal{S})| - \sum_{i=\{\mathtt{L},\mathtt{R}\}} \omega_i \log|\Gamma(\mathcal{S}_i)|, \quad \text{with} \quad \Gamma = \mathrm{diag}\left(\Lambda_1, \cdots, \Lambda_c, \cdots, \Lambda_{|\mathcal{C}|}\right). \quad (1)$$

Maximizing (1) encourages minimizing the determinant of the $6|\mathcal{C}| \times 6|\mathcal{C}|$ covariance matrix Γ, thus decreasing the uncertainty in the probabilistic vote cast by each cluster of voxels on each organ pose. Node growing stops when IG is below a fixed threshold, too few points reach the node or a maximum tree depth D is reached (here $D = 7$). After training, the j^{th} split node remains associated with the feature $\boldsymbol{\theta}_j$ and thresholds ξ_j, τ_j. At each leaf node we store the learned mean $\overline{\mathbf{d}}$ (with $\mathbf{d} = (\mathbf{d}_1, \cdots, \mathbf{d}_c, \cdots, \mathbf{d}_{|\mathcal{C}|})$) and covariance Γ, (fig. 4b).

This framework may be reformulated using non-parametric distributions, with pros and cons in terms of regularization and storage. We have found our parametric assumption not to be restrictive since the multi-modality of the input space is captured by our hierarchical piece-wise Gaussian model.

Discussion. Equation (1) is an information-theoretical way of maximizing the confidence of the desired continuous output *for all* organs, without going through intermediate voxel classification (as in [14] where positive and negative examples of organ centres are needed). Furthermore, this gain formulation enables testing different context models; *e.g.* imposing a *full* covariance Γ would allow correlations between all walls in all organs, with possible over-fitting consequences. On the other hand, assuming a *diagonal* Γ (and diagonal class covariances Λ_c) leads to uncorrelated output predictions. Interesting models live in the middle ground, where Γ is sparse but correlations between selected subgroups of classes are enabled, to capture *e.g.* class hierarchies or other forms of spatial context. Space restrictions do not permit a more detailed description of these issues.

Forest testing. Given a previously unseen CT volume $\mathcal{V}$, each voxel $\mathbf{v} \in \mathcal{V}$ is pushed through each tree starting at the root and the corresponding sequence of tests applied. The voxel stops when it reaches its leaf node $l(\mathbf{v})$, with l indexing leaves across the whole forest. The stored distribution $p(\mathbf{d}_c|l) = \mathcal{N}(\mathbf{d}_c; \overline{\mathbf{d}}_c, \Lambda_c)$ for class c also defines the posterior for the absolute bounding box position: $p(\mathbf{b}_c|l) = \mathcal{N}(\mathbf{b}_c; \overline{\mathbf{b}}_c, \Lambda_c)$, since $\overline{\mathbf{b}}_c(\mathbf{v}) = \hat{\mathbf{v}} - \overline{\mathbf{d}}_c(\mathbf{v})$. The posterior probability for $\mathbf{b}_c$ is now given by

$$p(\mathbf{b}_c) = \sum_{l \in \tilde{\mathcal{L}}} p(\mathbf{b}_c|l) p(l). \tag{2}$$

$\tilde{\mathcal{L}}$ is a subset of all forest leaves. Here we select $\tilde{\mathcal{L}}$ as the set of leaves which have the smallest uncertainty (for each class c) and contain 1% of all test voxels. Finally $p(l) = 1/|\tilde{\mathcal{L}}|$ if $l \in \tilde{\mathcal{L}}, 0$ otherwise. This is different from averaging the output of all trees (as done *e.g.* in [9,10]) as it uses the most confident leaves, independent from which tree in the forest they come from.

Anatomy detection. The organ c is declared present in the scan if $p(\mathbf{b}_c = \tilde{\mathbf{b}}_c) > \beta$, with $\beta = 0.5$.

Anatomy localization. The final prediction $\tilde{\mathbf{b}}_c$ for the absolute position of the c^{th} organ is given by the expectation $\tilde{\mathbf{b}}_c = \int_{\mathbf{b}_c} \mathbf{b}_c p(\mathbf{b}_c) d\mathbf{b}_c$.

3 Results, Comparisons and Validation

This section assesses the proposed algorithm in terms of its accuracy, runtime speed and memory efficiency; and compares it to state of the art techniques.

Accuracy in anatomy localization. Qualitative results on automatic anatomy localization within previously unseen, whole-body CT scans are shown in fig. 5.

Quantitative evaluation. Localization errors are shown in table 1. The algorithm is trained on 55 volumes and tested on the remaining 45. Errors are defined as absolute difference between predicted and true wall positions. The table aggregates results over all bounding box sides. Despite the large data variability we

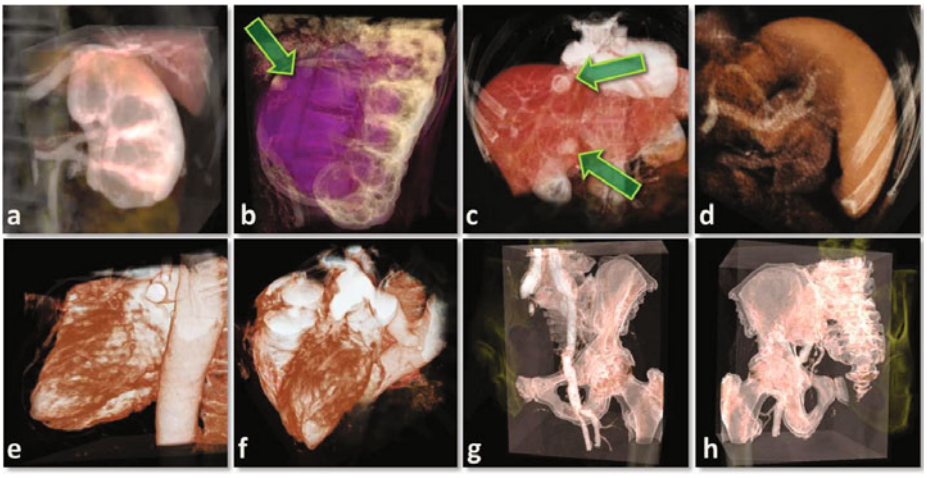


Fig. 5. Qualitative results showing the use of our automatic anatomy localizer for semantic visual navigation within 3D renderings of large CT studies. **(a)** The automatically computed bounding box for a healthy left kidney rendered in 3D. **(b)** As before but for a diseased kidney. **(c)** Automatically localized liver showing hemangiomas. **(d)** Automatically localized spleen, **(e, f)** heart and **(g, h)** left pelvis. Once each organ has been detected the 3D camera is repositioned, the appropriate cropping applied and the best colour transfer function automatically selected.

obtain a mean error of only $\sim 1.7cm$ (median $\sim 1.1cm$), sufficient to drive many applications. On average, errors along the z direction are about twice as large as those in x and y. This is due both to reduced resolution and larger variability in cropping along the z direction. Consistently good results are obtained for different choices of training set as well as different training runs.

Testing each tree on a typical 512^3 scan takes approximately $1s$ with our C++ implementation; and all trees are tested in parallel. Further speed-ups can be achieved with more low-level code optimizations.

Comparison with affine, atlas-based registration. One key aspect of our technique is its speed; important *e.g.* for clinical use. Thus, here we chose to compare our results with those obtained from a comparably fast atlas-based algorithm, one based on global registration. From the training set a reference atlas is selected as the volume which when registered with all *test* scans produced the minimum localization error. Registration was attained using the popular MedInria[3] package. We chose the global registration algorithm (from the many implemented) and associated parameters that produced best results on the *test* set. Such algorithm turned out to be block-matching with an affine transformation model. Note that optimizing the atlas selection and the registration algorithm on the test set produces results which are biased in favor of the atlas-based technique and yields a much tougher evaluation ground for our regression algorithm.

[3] `www-sop.inria.fr/asclepios/software/MedINRIA/`

Table 1. Regression forest results. Bounding box localization errors (in mm).

organ	heart	liver	spleen	left lung	right lung	left kidney	right kidney	gall bladder	left pelvis	right pelvis	*across all organs*
mean	15.4	17.1	20.7	17.0	15.6	17.3	18.5	18.5	13.2	12.8	**16.7**
std	15.5	16.5	22.8	17.2	16.3	16.5	18.0	13.2	14.0	13.9	**17.0**
median	9.3	13.2	12.9	11.3	10.6	12.8	12.3	14.8	8.8	8.4	**11.5**

Table 2. Atlas-based results. Bounding box localization errors (in mm).

organ	heart	liver	spleen	left lung	right lung	left kidney	right kidney	gall bladder	left pelvis	right pelvis	*across all organs*
mean	24.4	34.2	36.0	27.8	27.0	39.1	28.3	27.6	23.4	22.4	31.3
std	27.0	59.3	57.2	29.9	27.6	55.6	53.3	26.7	43.3	43.5	50.0
median	15.5	16.4	20.1	15.7	18.0	25.7	15.4	19.8	10.9	11.8	17.2

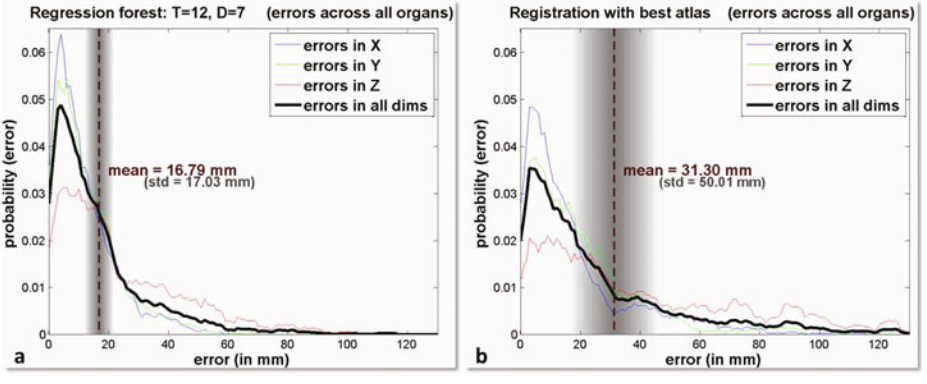


Fig. 6. Comparison with atlas-based registration. Distributions of localization errors for **(a)** our algorithm, and **(b)** the atlas-based technique. The atlas-induced errors show more mass in the tails, which is reflected by a larger standard deviation (std). The width of the vertical shaded band is proportional to the standard deviation.

The resulting errors (computed on the same test set) are reported in table 2. They show much larger error mean and standard deviation (about double) than our approach. Registration is achieved in between $90s$ and $180s$ per scan, on the same dual-core machine (*cf.* our algorithm runtime is $\sim 6s$ for $T = 12$ trees).

Figure 6 further illustrates the difference in accuracy between the two approaches. In the registration case larger tails of the error distribution suggest a less robust behavior[4]. This is reflected in larger values of the error mean and standard deviation and is consistent with our visual inspection of the registrations. In fact, in $\sim$ 30% cases the process got trapped in local minima and

[4] Because larger errors are produced more often than in our algorithm.

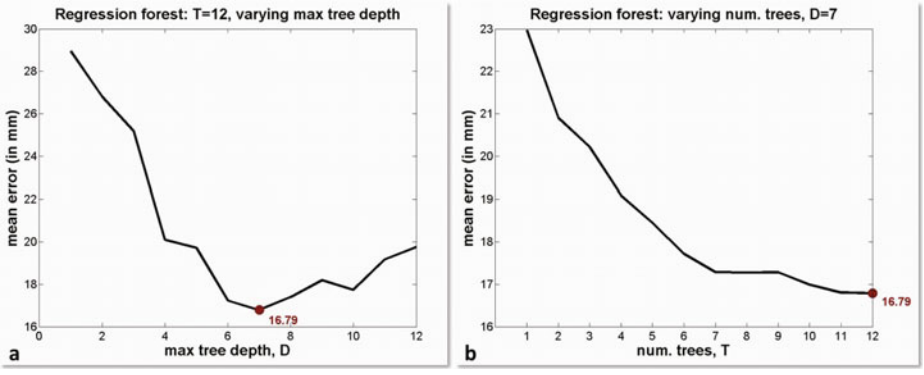


Fig. 7. Mean error as a function of forest parameters. (a) with varying maximum tree depth D and fixed forest size $T = 12$. **(b)** with fixed tree depth $D = 7$ and varying forest size T. All errors are computed on previously unseen test scans.

produced grossly inaccurate alignment. Those cases tend not to get improved when using a local registration step[5], while adding considerably to the runtime.

A regression forest with 6 trees takes $\sim$ 10MB of memory. This is in contrast with the roughly 100MB taken by each atlas. The issue of model size and runtime efficiency may be exacerbated by the use of more accurate and costly multi-atlas techniques [17]. Finally, in our algorithm increasing the training set usually decreases the test error without affecting the test runtime, while in multi-atlas techniques increasing the number of atlases linearly increases the runtime.

Comparison with voxel-wise classification. When compared to the classification approach in [14] we have found that our regression techniques produces errors less than half than those reported in [14] (on identical train and test sets) which, in turn demonstrated better accuracy than GMM and template-based approaches. In addition our regression algorithm computes the position of each wall (rather than just the organ centre), thus enabling approximate extent estimation.

Accuracy as function of forest parameters. Fig. 7 shows the effect of tree depth and forest size on accuracy. Trees deeper than 7 levels lead to over-fitting. This is not surprising as over-training with large trees has been reported in the literature. Also, as expected increasing the forest size T produces monotonic improvement without overfitting. No tree pruning has been employed here.

Automatic landmark detection. Fig 8 shows anatomical landmark regions automatically selected to aid organ localization. Given a trained tree and a chosen organ class (*e.g.* `left kidney`) we choose the two leaves with highest confidence. Then, we take the feature boxes (sect. 2.1) associated with the two closest ancestors (blue circles in fig. 8a) and overlay them (in green in fig. 8b,c) onto

[5] Which tends not to help escaping bad local minima.

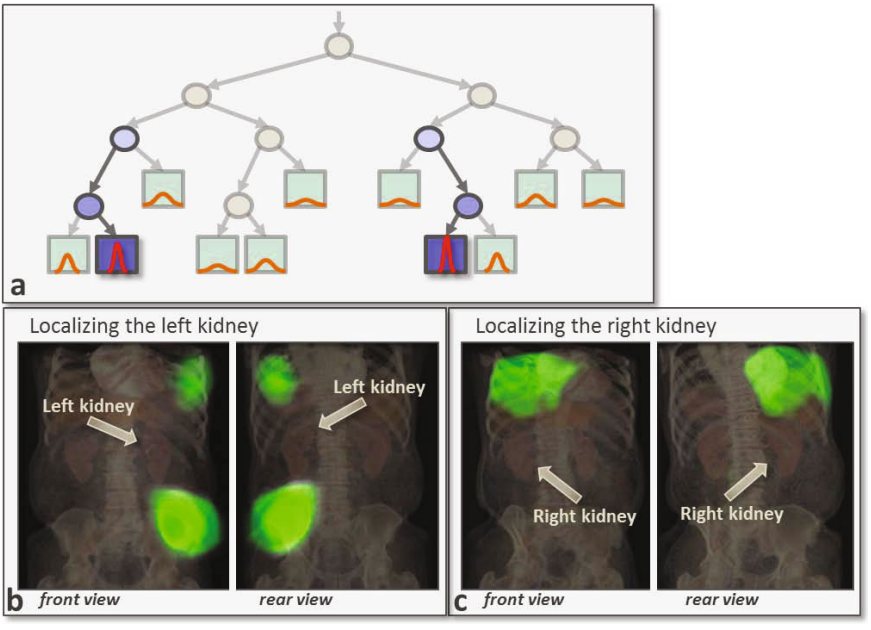


Fig. 8. Automatic discovery of salient anatomical landmark regions. (a) Given an organ class, the leaves associated to the two most confident distributions and two ancestor nodes are selected (in blue). **(b,c)** The corresponding feature boxes are overlayed (in green) on 3D renderings. The highlighted green regions correspond to anatomical structures which are automatically selected by the system to infer the position of the kidneys. See video in `http://research.microsoft.com/apps/pubs/default.aspx?id=135411`.

volumetric renderings, using the points reaching the leaves as reference. The green regions represent the anatomical locations which are used to estimate the location of the chosen organ. In this example the bottom of the left lung and the top of the left pelvis are used to predict the position of the left kidney. Similarly, the bottom of the right lung is used to localize the right kidney. Such regions correspond to meaningful, visually distinct, anatomical landmarks. They have been computed without any ground truth labels nor manual tagging.

4 Conclusion

Anatomy localization has been cast here as a non-linear regression problem where *all* voxels vote for the position of all anatomical structures. Location estimation is obtained via a multivariate regression forest algorithm which is shown to be more accurate and efficient than competing registration-based techniques.

At the core of the algorithm is a new information-theoretic metric for regression tree learning which enables maximizing the confidence of the predictions over the position of all organs of interest, simultaneously. Such strategy produces accurate predictions as well as meaningful anatomical landmark regions.

Accuracy and efficiency have been assessed on a database of 100 diverse CT studies. Future work includes exploration of different context models and extension to using other imaging modalities and non-parametric distributions.

References

1. Breiman, L.: Random forests. Technical Report TR567, UC Berkeley (1999)
2. Zhou, S.K., Zhou, J., Comaniciu, D.: A boosting regression approach to medical anatomy detection. In: IEEE CVPR, pp. 1–8 (2007)
3. Fenchel, M., Thesen, S., Schilling, A.: Automatic labeling of anatomical structures in MR fastView images using a statistical atlas. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 576–584. Springer, Heidelberg (2008)
4. Hardle, W.: Applied non-parametric regression. Cambridge University Press, Cambridge (1990)
5. Zhou, S., Georgescu, B., Zhou, X., Comaniciu, D.: Image-based regression using boosting method. In: ICCV (2005)
6. Friedman, J.: Greedy function approximation: A gradient boosting machine. The Annals of Statistics 2(28) (2001)
7. Vapnik, V.: The nature of statistical learning theory. Springer, Heidelberg (2000)
8. Yin, P., Criminisi, A., Essa, I., Winn, J.: Tree-based classifiers for bilayer video segmentation. In: CVPR (2007)
9. Montillo, A., Ling, H.: Age regression from faces using random forests. In: ICIP (2009)
10. Gall, J., Lempitsky, V.: Class-specific Hough forest for object detection. In: IEEE CVPR, Miami (2009)
11. Zhan, Y., Zhou, X.S., Peng, Z., Krishnan, A.: Active scheduling of organ detection and segmentation in whole-body medical images. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 313–321. Springer, Heidelberg (2008)
12. Torralba, A., Murphy, K.P., Freeman, W.T.: Sharing visual features for multiclass and multiview object detection. IEEE Trans. PAMI (2007)
13. Seifert, S., Barbu, A., Zhou, S.K., Liu, D., Feulner, J., Huber, M., Sühling, M., Cavallaro, A., Comaniciu, D.: Hierarchical parsing and semantic navigation of full body CT data. In: Pluim, J.P.W., Dawant, B.M. (eds.) SPIE (2009)
14. Criminisi, A., Shotton, J., Bucciarelli, S.: Decision forests with long-range spatial context for organ localization in CT volumes. In: MICCAI Workshop on Probabilistic Models for Medical Image Analysis (2009)
15. Shimizu, A., Ohno, R., Ikegami, T., Kobatake, H.: Multi-organ segmentation in three-dimensional abdominal CT images. Int. J. CARS 1 (2006)
16. Yao, C., Wada, T., Shimizu, A., Kobatake, H., Nawano, S.: Simultaneous location detection of multi-organ by atlas-guided eigen-organ method in volumetric medical images. Int. J. CARS 1 (2006)
17. Isgum, I., Staring, M., Rutten, A., Prokop, M., Viergever, M.A., van Ginneken, B.: Multi-atlas-based segmentation with local decision fusion—application to cardiac and aortic segmentation in ct scans. IEEE Trans. Medical Imaging 28(7) (2009)
18. Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A.: Classification and Regression Trees. Chapman and Hall/CRC (1984)
19. Shotton, J., Winn, J., Rother, C., Criminisi, A.: Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. In: IJCV (2009)

Correcting Misalignment of Automatic 3D Detection by Classification: Ileo-Cecal Valve False Positive Reduction in CT Colonography

Le Lu, Matthias Wolf, Jinbo Bi, and Marcos Salganicoff

CAD & Knowledge Solutions, Siemens Healthcare, Malvern, PA 19355, USA

Abstract. Ileo-Cecal Valve (ICV) is an important small soft organ which appears in human abdomen CT scans and connects colon and small intestine. Automated detection of ICV is of great clinical value for removing false positive (FP) findings in computer aided diagnosis (CAD) of colon cancers using CT colongraphy (CTC) [1,2,3]. However full 3D object detection, especially for small objects with large shape and pose variations as ICV, is very challenging. The final spatial detection accuracy often trades for robustness to find instances under variable conditions [4].

In this paper, we describe two significant post-parsing processes after the normal procedure of object (e.g., ICV) detection [4], to probabilistically interpret multiple hypotheses detections. It achieves nearly 300% performance improvement on (polyp detection) FP removal rate of [4], with about 1% extra computional overhead. First, a new multiple detection spatial-fusion method utilizes the initial single detection as an anchor identity and iteratively integrates other "trustful" detections by maximizing their spatial gains (if included) in a linkage. The ICV detection output is thus a set of N spatially connected boxes instead of a single box as top candidate, which allows to correct 3D detection misalignment inaccuracy. Next, we infer the spatial relationship between CAD generated polyp candidates and the detected ICV bounding boxes in 3D volume, and convert as a set of continuous valued, ICV-association features per candidate which allows further statistical analysis and classification for more rigorous false positive deduction in colon CAD.

Based on our annotated 116 training cases, the spatial coverage ratio between the new N-box ICV detection and annotation is improved by 13.0% (N=2) and 19.6% (N=3) respectively. An evaluation on large scale datasets of total $\sim$ 1400 CTC volumes, with different tagging preparations, reports average 5.1 FP candidates are removed at Candidate-Generation stage per scan; and the final CAD system mean FP rate drops from 2.2 to 1.82 per volume, without affecting the sensitivity.

1 Introduction

Colorectal cancer is the second leading, death-causing cancer for western population. Many computer aided diagnosis (CAD) systems [1,2,3,5] have been proposed to tackle the colonic polyp detection problem, with better accuracy and

B. Menze et al. (Eds.): MICCAI 2010 Workshop MCV, LNCS 6533, pp. 118–129, 2011.

sensitivity than radiologist alone. The most critical affecting factor for radiologists to accept the daily usage and adding value of a CAD system is its False Positive (FP) rate per scan or patient, while keeping high detection sensitivity. This is also the major difference from a good (helpful) to bad (misleading) CAD system [5]. Out of all FP types, Ileo-Cecal Valve has many (bumpy) polyp-like substructures which can confuse CAD algorithms and result as one of the most "difficult-to-remove" FP subgroup. As reported in the most recent study [6], 18.8% FPs are contributed by ICV structures, for a CAD system operating at 4.7 FPs per scan with reasonable sensitivity rate.

Detecting and segmenting small, soft and deformable human anatomic structures (e.g., Ileo-Cecal Valve) in a large 3D image volume (often > 500 slices) is a very challenging task. Ileo-Cecal Valve is highly deformable in shape and location by nature (without rigid attachment as connecting colon and small intestine), which leads to large intra-class shape, appearance and pose variations. In [4], we propose a generic object detection method to localize and segment an anatomic structure, such as Ileo-Cecal Valve in abdominal CT volumes, through an incremental parameter learning and registration procedure by sequentially aligning a bounding box with full 3D spatial configuration (i.e., 3D translation, 3D scaling and 3D orientation) towards the real structure. ICV has to be detected at the correct spatial scale range to understand its full context, and disambiguate from local, polyp-like subcomponents. The system diagram of ICV detection is shown in Fig. 1. For robustness, all steps of this detection pipeline leverage and keep multiple hypotheses (as a set of 3D boxes) for the next level until the last stage, which is in the same spirit of robust object tracking using multiple hypotheses [7], sequential Monte Carlo or particle filtering [8].

Exhaustive search of the 9-dimensional parameter space for the global optimal ICV bounding box is not only computationally infeasible, but also can not be trained due to the exponential-complexity negative class sampling issue in high dimensional parameter space. Though a high detection rate is achieved in [4], the spatial coverage ratios between computer detections and the annotated or desirable ICV bounding boxes are in need for improvement (probably inferior to face detection overlapping accuracy in 2D images due to higher dimensional parameter space of 9 versus 4). Especially for FP removal purpose in a CAD system, more spatially accurate detection of ICV leads to better reasoning of the spatial association between polyp, and ICV detections, which permits to remove more ICV false positives[1] [5,6].

In this paper, we present a sequence of significant post-parsing processes of [4], by spatially fusing the multiple ICV detection hypotheses in an "anchor-linking" fashion, constructing statistical features (e.g., distance, spatial-decaying

[1] For example, in our CAD system, overall $\approx 0.76\%$ FP candidates survive after the final classification, while as a specific FP category, the survival rate of ICV candidates is significantly higher as 7.83%. From the other viewpoint, ICV candidates form $\leq 1\%$ of the overall polyp candidates at CG level, but more than $10 \sim 15\%$ of the final system output FPs are composed of ICV (causing) FPs, if no explicite ICV candidates/FPs removal module is applied.

detection probabilities) continuously describing the underlying "candidate-ICV" associations, and building a discriminative classifier using new ICV features to remove false positives, while keeping the overall polyp detection sensitivity unchanged. The "anchor-linking" multiple detection fusion is related to component based object detection methods [9,10], but different in maximizing the trustful object region recovery by linking a few spatially correlated, "strong" detection candidates, while [9,10] aggregate multiple part-based detections to form the whole-object identification. The feature extraction and classification treatment from detection, enables more rigorous statistical analysis and removes about 90% more ICV type (polyp) FPs (0.38 versus 0.2 per volume, due to ICV existence) than improved N-box detection (n=3). Compared with [4], FP removal rate is nearly 300% (i.e., 0.38 versus 0.13 per scan). The computational overhead of post-parsing is neglectable compared with the ICV detection process [4].

2 Materials and Methods

In this section, we will first review the two-staged workflow of ICV detection by prior learning and incremental parameter learning [4]. Then a multiple detection fusion method, to improve the spatial coverage between the detected ICV area (a union of bounding boxes) and the true ICV occupying area, is described. Finally we map the spatial association between polyp detection candidates and the updated ICV detection output, to a set of four features including $\{Indicator_{ICV}, Prob_{ICV}, Dist_{ICV}, ProbDecay_{ICV}\}$ or $\{Indicator, Prob, Dist, ProbDecay\}$, by incorporating both localization (geometry) and detection (probability) information, and feed them into statistical analysis and classification for ICV-type FP reduction.

2.1 Progressive Ileo-Cecal Valve Detection in 3D

ICV detection is very challenging due to ICV's large variations in terms of its internal shape/appearance and external spatial configurations: $(X, Y, Z; S_x, S_y, S_z; \psi, \phi, \omega)$, or $(\Omega_T; \Omega_S; \Omega_R)$. To address these difficulties, we develop a two-staged approach that contains the prior learning to prune ICV's spatial configurations in position and orientation, followed by the position, size and orientation estimation of incremental parameter learning. The prior learning is inspired by the fact that, if likely hypotheses for ICV orifice can be found, its position in Ω_T can be constrained, then no explicitly exhaustive searching of position is needed. The ICV orifice has an informative, but not uniquely, distinctive surface profile that can possibly indicate ICV locations. It is also known that ICV orifice only lies on the colon surface that is computed using a 3D version of Canny edge detection. Thus we can prune all voxel locations inside the tissue or in the air for even faster scanning. Then given detected ICV orifice position from prior learning, we can further use this to constrain ICV's location for efficient scanning, as described in [4].

Fig. 1 shows the diagram of our full detection system of two stages and five individual steps. Each step of encoding process is formulated as a general binary classification problem, and is specifically implemented using probabilistic

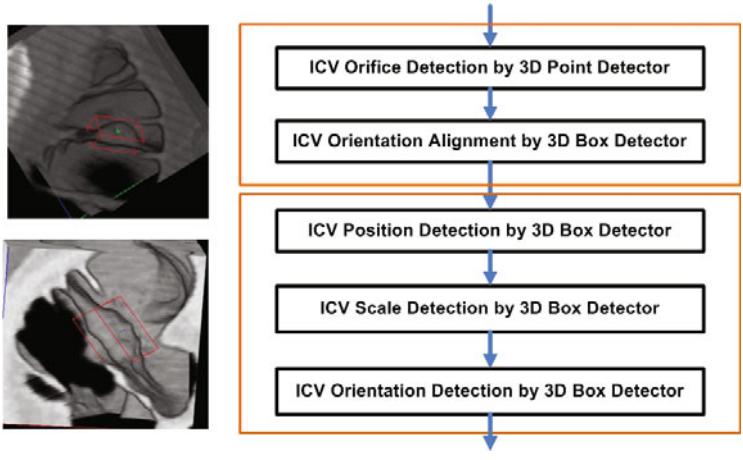


Fig. 1. System diagram of Ileo-Cecal Valve detection, with prior learning (upper block) and incremental parameter learning (lower block)

boosting tree algorithm (PBT) [11]. To learn the object (e.g., ICV) appearance model, we employ 3D steerable features [12] which are composed by a number of sampling grids/points where 71 local intensity, gradient and curvature based features are computed at each grid. The whole sampling pattern models semi-local context. In contrast to popular 3D HAAR features [13], only the sampling grid-pattern of steerable features need to be translated, rotated and re-scaled instead of data volumes. It allows fast 3D data evaluation and has shown to be effective for object detection tasks [12]. The separation of sampling grid pattern and local 71 gradient/curvature features allows the flexibility of different geometric structure designs of grid-pattern as spatial assembles of unchanged local features. Particularly, an axis-based pattern is proposed for detecting ICV's orifice at step 1, and a box-based pattern for parsing the ICV orientation, scale and size at following steps, with total 5751 or 52185 local features for boosting respectively.

If there is only one existing object per volume (such as ICV) and the training function can be perfectly learned by a classifier at each step, setting only one detection candidate (e.g., $M = 1$) per step is sufficient to achieve the correct detection. In practice, we set $M = 50 \sim 100$ for all intermediate detection steps to improve robustness. It means that we maintain multiple detection hypotheses until the final result. For the description of how training parameters are obtained in this multi-stage detection hierarchy, refer to [4] for details.

Improvements: ICV contains many polyp-like local structures which often survive through colon CAD systems. By localizing a spatially accurate bounding box of ICV, this type of ambiguous false positives as generated by an initial candidate-generation (CG) process (within the above detected bounding box), can be removed. For this task, 1), we further enhanced the ICV orifice detection

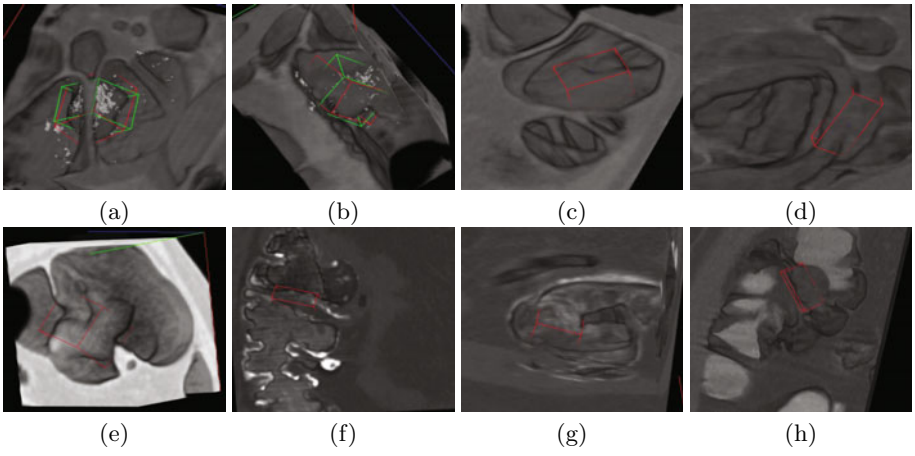


Fig. 2. (a,b) An example of ICV detection result from two viewpoints. The red box is the annotation; the green box is the detection. (c,d,e,f,g,h) Examples of ICV detection results from unseen clean colon CT volumes (c,d,e) and unseen solid (f) or liquid tagged (g,h) colon CT volumes.The red box is the final detection result where no annotation available. Note that only a CT subvolume surrounding the detected ICV box is visualized for clarity. This picture is better visualized in color.

stage (as the first step in Fig. 1) by adding all labeled polyp surface voxels into its negative training dataset, which results a **propose-specific** and **more discriminative** training against losing polyps or reducing sensitivity. Other stages are consequentially retained in the same way. 2), **Non-Maximum suppression** is also performed after the prior learning by only keeping the top ICV box candidate at each different location. This further increases the spatial sampling and computational efficiency, as more spatial regions will be exploited by later training and classification stages with the same computational budget, or the number of kept samples. Some positive ICV detections are illustrated in Fig. 2. The processing time varies from $4 \sim 10$ seconds per volume on a P4 3.2G machine with 2GB memory.

2.2 Contextual N-Box ICV Detection by Spatial Fusion

To obtain a more precise 3D ICV region from detection, a contextual N-box model is employed. 1), We use the single ICV detection box B_1 as an anchor to explore other reliable expansions. The trust or reliability is guaranteed by maintaining other boxes with both posterior probabilities above a high threshold and good overlaps with the anchor box. For all other hypotheses $\{\hat{B}_i\}$ (except B_1) returned in the last step of detection, we first apply a prefilter and only retain "trustful" candidates satisfying $\gamma(B_1, \hat{B}_i) \geq \gamma_1$ and $\rho(\hat{B}_i) \geq \rho_1$ where $\gamma(\bullet, \bullet)$ computes the spatial overlap ratio between two boxes and $\rho(\hat{B}_i)$ returns the posterior detection probability of $\hat{B}_i$ from above [4]. The two constraints guarantee that B_2 is spatially correlated with B_1 ($\gamma_1 = 0.5$) and is a high

quality ICV detection by itself $\rho_1 = 0.8$. 2), Then we sort them according to their spatial gains $Vol(\hat{B}_i - B_1 \bigcap \hat{B}_i)$ and the box that gives the largest gain is selected as the second box B_2. Our boxes are fully mathematically parameterized which allows fast evaluation of voxel overlapping. 3), By taking B_1 and B_2 as a union $Box_d = B_1 \bigcup B_2$, it is straightforward to expand the model for N-box ICV model with $N > 2$, by maximizing $Vol(\hat{B}_i - Box_d \bigcap \hat{B}_i)$. The union Box_d grows by adding one new winning box per iteration. Let Box_a be the annotated bounding box of the Ileo-Cecal Valve and Box_d be the detected N-Box. The spatial overlap ratio between Box_a and Box_d is defined as

$$\gamma(Box_a, Box_d) = \frac{Vol(Box_a \bigcap Box_d)}{Vol(Box_a) \bigcup Vol(Box_d)} \tag{1}$$

where $Vol()$ is the box-volume function (eg. the voxel number inside a box). The spatial coverage ratio of Box_d and Box_a is defined as

$$\alpha(Box_a, Box_d) = \frac{Vol(Box_a \bigcap Box_d)}{Vol(Box_a)} \tag{2}$$

which describes the percentage of the annotated ICV area covered by the detection Box_d. In practice, the number N of boxes in Box_d can be determined by cross-validation, by maximizing $\alpha(Box_a, Box_d)$ under the constraint of maintaining $\gamma(Box_a, Box_d)$ at high level. $\alpha(Box_a, Box_d)$ is the direct performance measure, as the percentage of the true ICV volumetric region Box_a recovered by Box_d, which impacts on the ratio of ICV causing FPs in Box_a that can be removed by Box_d instead. High overlap ratio $\gamma(Box_a, Box_d)$ as *Jaccard similarity*, keeps detection Box_d highly confident against ground truth Box_a. A balance between $\alpha(Box_a, Box_d)$ and $\gamma(Box_a, Box_d)$ needs to be achieved. A few illustrative examples of multi-box ICV Detection are shown in Fig. 3.

2.3 Features and ICV False Positive Classification

Given the ICV detection output $Box_d = \{B_i\}_{i=1,2,...,N}$ and the spatial locations $\{L_j\}$ of a set of polyp candidates (in the order of hundreds per volume), we first compute the Euclidean distances (Geodesic distance is more desirable but with higher computational cost; Euclidean is a close approximation in low distance range) for each of polyp candidate against Box_d as

$$Dist(L_j, Box_d) = \min_i Dist(L_j, B_i) \tag{3}$$

Then, the generic "point-to-box" distance in 3D is converted as a standard "point-to-triangle" distance because the box is a spatially convex set including all voxels inside. The triangle is found by selecting the set of three box vertices $B_{i,k}, k = 1, 2, 3$ with the smallest "point-to-point" Euclidean distances $\parallel B_{i,k} - L_j \parallel$ according to L_j. Thus $B_{i,k}, k = 1, 2, 3$ may vary against different L_j. Finally the "point-to-triangle" distance is calculated using the standard geometric algorithm [14] and we denote $Dist(L_j, Box_d)$ as $Dist^j_{ICV}$ for any jth polyp

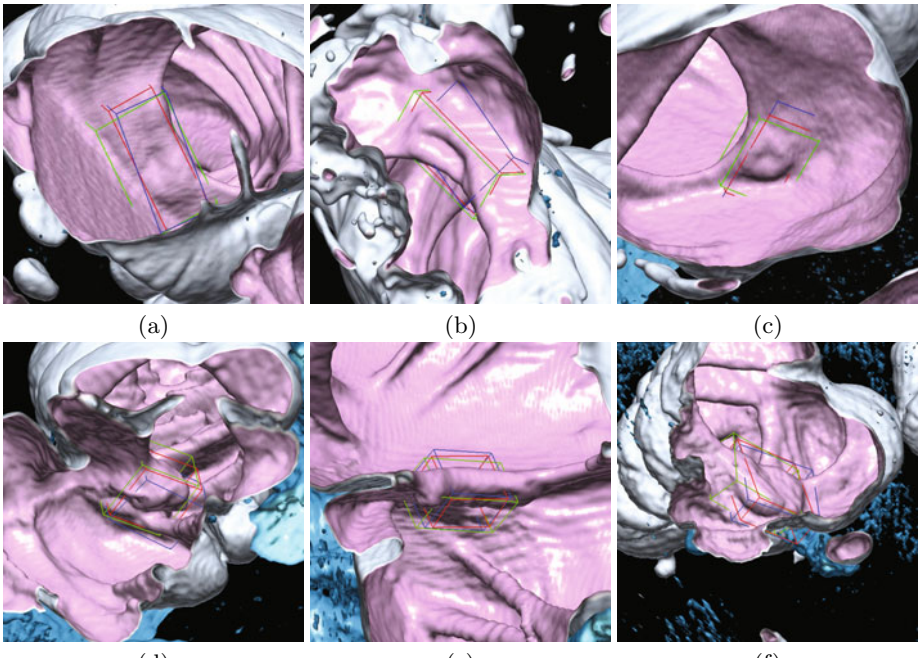


Fig. 3. Multi-box ICV Detection results (N=3) with clean preparation (a,b,c) and tagged preparation (d,e,f). Note that rugged surface is more visible in (d,e,f) under tagged preparation which potentially causes more challenges for ICV detection or degrading on localization accuracy. Notice that the spatial coverage of ICV detection boxes against the true ICV area improves as N increases from 1 to 2, 3. The first, second, and third detection box is color-coded as red, green, and blue respectively. This picture is better visualized in color.

candidate. Note that if L_j is determined inside any box $\subset Box_d$, $Dist^j = 0$ will be automatically set, without any distance calculations. Furthermore, a binary indicator $\{Indicator^j_{ICV}\}$ is also derived from $Dist^j_{ICV}$

$$Indicator^j = \begin{cases} True, \ if Dist^j = 0; \\ False, \ \ otherwise. \end{cases} \tag{4}$$

The confidence of ICV detection procedure can also be explored as $Prob_{ICV}$ volumewise, regardless of different CG candidates. Lastly, by combining the information of the overall detection probability $Prob_{ICV}$ per volume (only one ICV per abdominal scan) and $\{Dist^j_{ICV}\}$ per candidate, a new hybrid feature $ProbDecay^j_{ICV}$ is computed as

$$ProbDecay^j = Prob \times exp(-Dist^j/\sigma) \tag{5}$$

It simulates the spatially decaying effect of ICV detection probability $Prob_{ICV}$ propagating from Box_d to the location L_j where σ controls the decaying speed

factor and is determined by cross-validation, or multiple σ can be employed for decaying with respect to different spatial scales. $ProbDecay^j$ integrates the cues of distance $Dist^j$, detection posterior probability $Prob$ and the spatial scale σ, where σ can be set by optimizing $ProbDecay^j$'s classification performance (e.g., Fisher score [15]). As seen later, this feature demonstrates the best effectiveness on modeling the relationship or association between polyp candidates and the detected ICV, and removing ICV type FPs via classification, out of four features. Using geodesic distance to replace $Dist^j$ is probably more sensible and accurate for modeling the confidence propagation over surface, because all anatomical structures (e.g., ICV, polyp), being interested for CTC lie on colonic surface, and surface geodesic coordinates normally serve as their spatial locations. This is left for future work.

In summary, we obtain a set of four features $\{Indicator^j, Prob, Dist^j, ProbDecay^j\}$ for any jth polyp candidate, and these features can be used to train a classifier to report whether a candidate is truly associated with ICV rather than a polyp. Of course, these four features are not statistically independent, but in section 3, their joint discriminative capability is shown to be higher than each individual, and thus is finally used for the best classification performance on filtering out ICV type FPs, using Quadratic/Linear Discriminant Analysis classifiers [15].

Previous Work: The closest previous work is by Summer et al. [16,3] which however is drastically different from ours in two aspects. (1) For localization of ICV, [3] relies on a radiologist to interactively identify the ICV by clicking on a voxel inside (approximately in the center of) the ICV, as a requisite, manual initialization step, followed by classification process. (2) For classification, some human designed heuristic rules based on ICV volume and attenuation thresholds are utilized. Refer to [16,3] for details. Their overall sensitivity of ICV detection is 49% and 50% based on the testing (70 ICVs) and training datasets (34 ICVs) [3], respectively.

3 Experimental Results

Detection Performance: Our ICV detection process is trained with an annotation dataset of 116 volumes (clean-prep), where each ICV per volume is precisely bounded using a 3D box with nine degrees of freedom (3D location, orientation and scale) by two experts, as shown in Fig. 1. 1), Our initial experimental assessment in training shows that 2-box model improves the mean coverage ratio $\alpha(Box_a, Box_d)$ from 75.6% to 88.6%. When $N = 3$, the $\alpha(Box_a, Box_d)$ reaches 95.2%. $\gamma(Box_a, Box_d)$ are 72.7%, 85.2%, 86.1% for $N = 1, 2, 3$ and starts to decrease slightly for $N > 3$. Finally, N is chosen to be 3. 2), For testing cases where there are no annotation of ICV bounding boxes available, hence we evaluate the ICV detection rates by inspecting each "anchor" box returned by our ICV detector, and labeling it as *true* or *false*, using two unseen testing datasets of 526 volumes (clean-prep) and 689 volumes (fecal tagging-prep including both iodine and barium preparations) respectively. Siemens, GE and Philips scanners are

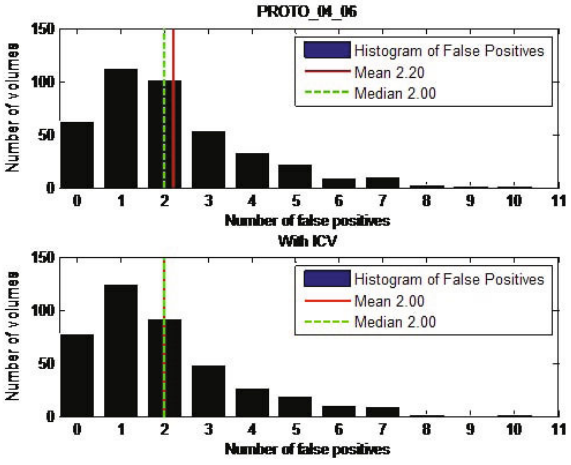


Fig. 4. Volume-FP count histograms of tagging testing dataset, before (**UPPER**) and after (**LOWER**) N-box ICV post-filter processing

used for image acquisition, under different imaging protocols, from 10+ medical sites in Asia, Europe and USA. The detection rates are 91.3% and 93.2% for clean and tagged datasets.

False Positive Detection: FP deduction is also tested on our clean and tagged training/testing datasets. There is no significant statistical performance difference among different datasets, and the detailed analysis and results on tagged testing dataset are reported below. The ICV detection can be implemented as both pre-filter and post-filter for our existing CTC CAD system. In post-processing, only those candidates that are labeled as "Polyp" in the final classification phase are used for screening; while as pre-filter, all candidates output by an initial Candidate-Generation (CG) scheme are employed. With N-box ICV detection improvement, the final number of false positives drops from 2.2 fp/volume to 2.0 fp/volume (removing FPs with $Dist^j = 0$ or $Indicator^j = true$, 90% improvement), without reducing the overall sensitivity of the CAD system. For $N = 1$, the detection based removal [4] in average rejects 0.13 (or 5.91%) FPs per volume. When used as ICV pre-filter, the average FP removal is 3.1 per volume. Compared with $N = 1$ in [4], multiple-box ($N = 3$) ICV FP classification process removes 62.5% more CG candidates for this dataset. FP histogram of the tagged testing dataset of 412 volumes, before and after ICV post-filter, demonstrates the advantageous performance impact of using multi-detection fusion, as shown in Fig. 4. The lower histogram has more mass moving towards the left (as smaller FP numbers).

False Positive Classification: We first evaluate Fisher Discriminant Scores (FD) of any continuous valued feature $f \in \{Dist^j, ProbDecay^j\}$ over CG candidates, defined as

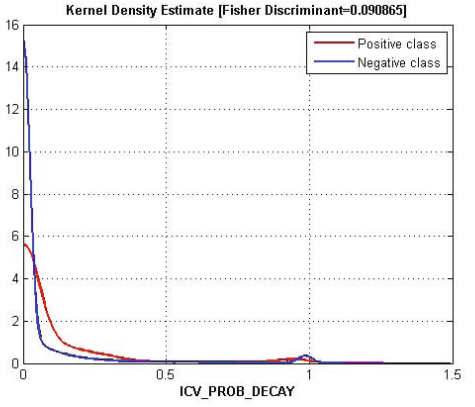


Fig. 5. Kernel Density Estimate plots of the spatial-probability feature *ProbDecay* for positive (polyp) and negative (non-polyp) classes, with the fisher score 0.0909

$$J(f) = \frac{(\bar{f}^+ - \bar{f}^-)^2}{\sigma^2(f^+) + \sigma^2(f^-)} \tag{6}$$

where $\bar{f}^+$ and $\bar{f}^-$ denote the mean; $\sigma^2(f^+)$ and $\sigma^2(f^-)$ represent the covariance of f distribution on positive $\{f^+\}$ (polyp) and negative $\{f^-\}$ (non-polyp) classes. $ProbDecay^j$ ($\sigma = 10mm$) returns higher FD score[2] of 0.0909 than $Dist^j$. The feature's Kernel Density Estimate plots are drawn in Fig. 5. This means that the hybrid feature $ProbDecay_{ICV}$ can describe underlying soft "candidate-ICV affiliations" more precisely and is probably more effective on removing more ICV related FPs, via inferring both spatial and detection probability information. Next, we train a *Linear Discriminant Classifier* on $\{Prob, Dist^j, \{ProbDecay^j_{\sigma=5,10,15,20}\}\}$ of all candidates using tagged training dataset and obtain the projection $\{\phi^j\}$ as a new "summarization" feature, which indeed has a better FD score of 0.171 and $\sigma = 5, 10, 15, 20$ simulates the multiscale effect of $ProbDecay^j$. $\{\phi^j\}$ is further thresholded for recognizing ICV type FPs from polyp candidates. Based on this, we report an average of 5.1 false positives removed per volume at CG stage; and the final CAD system FP rate also decreases from 2.2 to 1.82 per volume (or, 17.2% of all FPs), for tagged testing dataset. The sensitivities remain the same at both stages. Compared with the results of binary decision $Dist^j = 0$ with $N = 3$, the performance improvements of leveraging the continuous feature ϕ^j, are 64.5% and 90.6% at CG or final system level, respectively. In [6], 18.8% of 4.7 FPs is caused by ICV which is 0.87 per volume. The numerical results of FP histograms in tagged testing dataset, without and with the classification ICV post-filter using $\{\phi^j\}$, are given in Table 1. This improves our previous result of *False Positive Detection* as in Fig. 4. It is clearly noticeable more volumes have even lower (per-volume) FP rates. From

[2] Since the majorities of both positive and negative distributions are out of the realm of ICV spatial occupations, the absolute FD scores are not very high generally.

Table 1. Volume-FP count histograms of tagging testing dataset, without and with enhenced ICV post-filter on $\{\phi^j\}$

False Positive Histogram	Without ICV Filter	With ICV Filter
0	62 [15.05%]	90 [21.84 %]
1	113 [27.43%]	129 [31.31 %]
2	102 [24.76%]	92 [22.33 %]
3	54 [13.11%]	40 [9.71 %]
4	33 [8.01%]	22 [5.34 %]
5	22 [5.34%]	17 [4.13 %]
6	9 [2.18%]	11 [2.67%]
7	10 [2.43%]	8 [1.94%]
8	3 [0.73%]	1 [0.24%]
9	2 [0.49%]	2 [0.49%]
≥ 10	2 [0.49%]	1 [0.24%]

our further evaluation, this improvement also generalizes well to clean training and testing datasets, with similar observation obtained. As a post-processing, the additional computation expense over [4] is approximately 1%.

4 Discussion

In this paper, we propose a sequential "anchor-linking" approach on multiple detection hypotheses, to improve the alignment accuracy of automatic 3D detection for Ileo-Cecal Valve. The final ICV detection output is a set of spatially connected N-boxes where our method is generic and applicable to other 3D/2D multi-hypothesis detection problems [9,10,7,8], without restricting to [4]. Then we derive continuous valued features (e.g., $\{Dist^j, ProbDecay^j\}$) more precisely describing the underlying "candidate-ICV" associations, which permits further statistical analysis and classification, converting from binary detections. Significant performance improvement is demonstrated on ICV-relevant false positive reduction rates in CT Colonography, compared with previous work [4,16,3], without sacrificing polyp detection sensitivity.

References

1. Yoshida, H., Dachman, A.H.: Cad techniques, challenges, and controversies in computed tomographic colonography. Abdominal Imaging 30, 26–41 (2005)
2. Bogoni, L., Cathier, P., et al.: Cad for colonography: A tool to address a growing need. The British Journal of Radiology 78, 57–62 (2005)
3. O'Connor, S., Summers, R., Yao, J., et al.: Ct colonography with computer-aided polyp detection: volume and attenuation thresholds to reduce false-positive findings owing to the ileocecal valve. Radiology 241, 426–432 (2006)

4. Lu, L., Barbu, A., Wolf, M., Liang, J., Bogoni, L., Salganicoff, M., Comaniciu, D.: Simultaneous detection and registration for ileo-cecal valve detection in 3D CT colonography. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part IV. LNCS, vol. 5305, pp. 465–478. Springer, Heidelberg (2008)
5. Park, H., et al.: Computer-aided polyp detection on ct colonography: Comparison of commercially and academically available systems. In: RSNA (2007)
6. Slabaugh, G., Yang, X., Ye, X., Boyes, R., Beddoe, G.: A robust and fast system for ctc computer-aided detection of colorectal lesions. Algorithms: special issue on Machine Learning for Medical Imaging 3(1), 21–43 (2010)
7. Cox, I.J., Hingorani, S.L.: An efficient implementation of reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking. IEEE Trans. on PAMI 18(2), 138–150 (1996)
8. Isard, M., Blake, A.: Condensation - conditional density propagation for visual tracking. International Journal of Computer Vision 29(1), 5–28 (1998)
9. Zheng, Y., Lu, X., et al.: Robust object detection using marginal space learning and ranking-based multi-detector aggregation: Application to left ventricle detection in 2d mri images. In: IEEE Conf. on CVPR (2009)
10. Wu, B., Nevatia, R., Li, Y.: Segmentation of multiple, partially occluded objects by grouping, merging, assigning part detection responses. In: IEEE CVPR (2008)
11. Tu, Z.: Probabilistic boosting-tree: Learning discriminative methods for classification, recognition, and clustering. In: IEEE ICCV (2005)
12. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Fast automatic heart chamber segmentation from 3d ct data using marginal space learning and steerable features. In: IEEE ICCV (2007)
13. Tu, Z., Zhou, X.S., Barbu, A., Bogoni, L., Comaniciu, D.: Probabilistic 3d polyp detection in ct images: The role of sample alignment. In: IEEE CVPR (2006)
14. Eberly, D.: 3d game engine design: A practical approach to real-time computer graphics, 2nd edn. Morgan Kaufmann, San Francisco (2000)
15. Duda, R., Hart, P., Stork, D.: Pattern Classification, 2nd edn. Wiley Interscience, Hoboken (2000)
16. Summers, R., Yao, J., Johnson, C.: Ct colonography with computer-aided detection: Automated recognition of ileocecal valve to reduce number of false-positive detections. Radiology 233, 266–272 (2004)

Learning Adaptive and Sparse Representations of Medical Images

Alessandra Staglianò, Gabriele Chiusano, Curzio Basso, and Matteo Santoro

DISI, Università di Genova,
Via Dodecaneso, 35
16146 Genova, Italy
{stagliano,chiusano,basso,santoro}@disi.unige.it

Abstract. In this paper we discuss the impact of using algorithms for dictionary learning to build adaptive and sparse representations of medical images. The effectiveness of coding data as sparse linear combinations of the elements of an over-complete dictionary is well assessed in the medical context. Confirming what has been observed for natural images, we show the benefits of using adaptive dictionaries, directly learned from a set of training images, that better capture the distribution of the data. The experiments focus on the specific task of image denoising and produce clear evidence of the benefits obtained with the proposed approach.

1 Introduction

Over the last decade, the research field of natural image analysis has witnessed a significant progress thanks to a more mature and effective use of algorithmic approaches to data representation and knowledge extraction. Among the most successful techniques, the ones based on *adaptive sparse coding* are playing a leading role. In this paper, we explore the possible application of such methods to the context of medical image analysis (MIA).

The insightful idea behind sparse coding is to build succinct representations of visual stimuli, which may be conveniently obtained in practice by decomposing the signals into a linear combination of a few elements from a given dictionary of basic stimuli, or *atoms* [6]. Adaptiveness, in this context, is achieved through learning the atoms directly from the data, a problem called *dictionary learning*, instead of using fixed over-complete dictionaries derived analytically, as it is the case for the popular wavelet-based approaches [13]. Significant experimental successes of adaptive sparse coding have been reported in different applications of computer vision, such as denoising [5], compression [1], texture analysis [15], scene categorization and object recognition (interesting examples are in [16], [12], or [9]).

The observation that motivates the present work is that the exploitation of these advanced techniques for dictionary learning and sparse coding in the field of MIA remains – to a large extent – unattained. To the best of our knowledge, [18] is the only work presenting results on medical images. Therefore, the first underlying goal of the paper is to promote a discussion on the potentials of

B. Menze et al. (Eds.): MICCAI 2010 Workshop MCV, LNCS 6533, pp. 130–140, 2011.

adaptive sparse coding for addressing different problems of interest to the audience of the workshop. In order to make the discussion concrete, we focus on two state-of-the-art dictionary learning algorithms. Specifically, we consider the K-SVD algorithm, introduced by Aharon et al. [1], and the method of Lee et al. [10], which in the rest of the paper we refer to as ℓ_1-regularized Dictionary Learning (ℓ_1-DL). After a preliminary description of their main common properties we provide details on the two methods and compare their performances using a collection of MR images representing different anatomical regions. In order to make more clear the methodological and algorithmic issues discussed in the following sections and to support experimentally some of the conclusions we have drawn in our work, we opted to focus on one specific task only, namely image denoising. In this context, by using the two learned dictionaries we succeeded in improving the results of a state-of-the-art method for denoising based on the Discrete Cosine Transform [8]. We also stress that denoising is only one of the possible applications of these techniques, and that higher-level tasks such as detection or classification are expected to benefit from them.

The paper is organized as follows. In the next sub-section we briefly review the main results in the literature on dictionary learning and connect the few papers on medical image analysis using these tools. In section 2 we provide a formal description of the problem, and an implementation-focused overview of K-SVD and ℓ_1-DL. Next, we present the experimental setting and the results obtained. We conclude with a brief discussion.

1.1 Relevant Literature

The use of fixed overcomplete representations, such as wavelets, to process and interpret medical images is already widespread and successful. Indeed, overcomplete dictionaries coupled with sparse coding have been shown in the past to yield more expressive representations, capable of achieving better performances in a wide range of tasks. However, we believe that further major advancements may be obtained by investigating methods for learning such representation from the data.

This problem, in its non-overcomplete form, has been studied in depth. Indeed, although not yielding overcomplete dictionaries, Principal Component Analysis (PCA) and its derivatives are at the root of such approaches, based on the minimization of the error in reconstructing the training data as a linear combination of the basis elements (see e.g. [3]).

The seminal work of Olshausen and Field [14] was the first to propose an algorithm for learning an overcomplete dictionary in the field of natural image analysis. Probabilistic assumptions on the data led to a cost function made up of a reconstruction error and a sparse prior on the coefficients, and the minimization was performed by alternating optimizations with respect to the coefficients and to the dictionary. Most subsequent methods are based on this alternating scheme of optimization, with the main differences being the specific techniques used to induce a sparse representation. Recent advances in compressed sensing and feature selection led to use ℓ_0 or ℓ_1 penalties on the coefficients, as in [1] and

[11], respectively. The former method has been applied to denoising in what is, to the best of our knowledge, the only paper applying such methods to medical images [18].

In [17], the authors proposed to learn a pair of encoding and decoding transformations for efficient representation of natural images. In this case the encodings of the training data are dense, and sparsity is introduced by a further non-linear transformation between the encoding and decoding modules.

Bag-of-features representations can be seen as a nearest-neighbor approximation of representations based on dictionary learning and sparse coding (see [20,1]). In [2] the authors tackle the problem of detecting lesions in cervigram images by employing such an approach. Image patches are represented by the closest codewords from a dictionary that has been learned by collecting patches from the training set, reducing their dimensionality by PCA, and selecting the codewords by k-means.

In summary, although a number of works have been published demonstrating the effectiveness of adaptive and sparse representations for natural image analysis, our paper is the first attempt to explicitly tackle this topic in the context of medical image analysis.

2 Learning Sparse Overcomplete Representations

In this section we briefly introduce some common properties of the algorithms for dictionary learning and their application to sparse image representation. We start by setting up the notation and recalling the mathematical properties of a general equation (see Eq. 2 below) from which the algorithms may be derived.

Let $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]$ be a $d \times N$ matrix whose columns are the training vectors. The objective of dictionary learning is to find a suitable $d \times K$ matrix $\boldsymbol{D}$, the *dictionary*, that allows to represent the training data as linear combinations of its columns $\boldsymbol{d}_j$, also called *atoms*. Denoting by $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N]$ the $K \times N$ matrix whose columns are the coefficients of the linear combinations, also known as *encodings* or *codes*, the problem can be formulated as the minimization of the reconstruction error

$$\min_{\boldsymbol{D},\boldsymbol{U}} \|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{U}\|_F^2. \tag{1}$$

It can be readily seen that, for $K \leq d$, viable methods to compute the minimizer of such equation are Principal Component Analysis (PCA) and its derivatives. In particular, among possibly different solutions of (1), PCA picks the one with minimal $\|\boldsymbol{U}\|_F^2$. For the reasons exposed in the introduction we are interested in over-complete settings, where $K > d$, with sparse representations, that is when the vectors $\boldsymbol{u}_i$ have few coefficients different from zero. This can be achieved by adding to the cost function (1) a penalty term (or a constraint) favoring the sparsity of $\boldsymbol{U}$:

$$\min_{\boldsymbol{D},\boldsymbol{U}} \|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{U}\|_F^2 + \tau P(\boldsymbol{U}). \tag{2}$$

Two possible choices for the penalty $P(\boldsymbol{U})$ have been explored in the literature. The most natural choice is probably one based on the ℓ_0 norm of the encodings $\boldsymbol{u}_i$, which counts the number of non-zero elements, but it is also the most difficult to optimize, leading to a non-convex problem. An alternative is the ℓ_1 norm, which can be shown to be equivalent to the former norm under appropriate conditions on $\boldsymbol{D}$, while being more easily tractable.

Almost all the proposed methods so far for learning the dictionary share a basic scheme, in which the problem is solved by alternating between two separate optimizations, with respect to $\boldsymbol{U}$ and $\boldsymbol{D}$. The former, where $\boldsymbol{D}$ is kept fixed, is the step of *sparse coding*, and is the only one directly affected by the choice of the penalty. There is now a wide literature on methods for solving the problem with different kinds of norm and we refer the interested reader to [4] for an in-depth treatment. The latter optimization step, with $\boldsymbol{U}$ fixed, performs a *dictionary update*. In principle this is an unconstrained Least Squares (LS) problem that could be solved in closed-form, as it is done by the *method of optimal directions* (MOD) [7]. However, choosing the ℓ_1 norm as sparsity penalty forces the introduction of a constraint on the norm of the atoms, leading to a quadratic-constrained LS.

Among the methods available in the literature for learning the dictionary through the minimization of some variant of equation (2) we focused specifically on two experimentally successful and widely used algorithms, described in the remainder of the section. Implementations of both algorithms are publicly available, and we believe this may be beneficial to prospective practitioners of dictionary learning for medical image analysis.

K-SVD algorithm. The first algorithm we reviewed is named K-SVD because its dictionary update step is essentially based on solving K times a singular value decomposition problem. K-SVD has been proposed in [1], and it aims at minimizing a variant of the functional (2) in which the sparsity is enforced by a constraint on the ℓ_0 norm of the columns of $\boldsymbol{U}$:

$$\min_{\boldsymbol{D},\boldsymbol{U}} \|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{U}\|_F^2 \quad s.t. \quad \|\boldsymbol{u}_i\|_0 \leq T_0. \tag{3}$$

Among the nice features of the K-SVD we deem worth mentioning are: (i) the dictionary update is performed efficiently atom-by-atom; (ii) the iterations are shown empirically to be accelerated by updating both the current atom and its associated sparse coefficients simultaneously. Since the algorithm uses an ℓ_0 constraint, in our experiments we used Orthogonal Matching Pursuit (OMP) for solving the sparse coding step [19].

ℓ_1-DL Algorithm. The second algorithm we considered was proposed by Lee and colleagues in [10] and aims at minimizing the functional (2) with the standard ℓ_1 penalty:

$$\min_{\boldsymbol{D},\boldsymbol{U}} \|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{U}\|_F^2 + \tau \sum \|\boldsymbol{u}_i\|_1 \quad s.t. \quad \|\boldsymbol{d}_i\|_2 \leq 1 \tag{4}$$

As pointed out above, in this case adding a further constraint on the norm of the atoms is needed to avoid degenerate solutions where the norm of the coefficients is kept small by scaling up the atoms.

The algorithm is based on the alternating convex optimization over $\boldsymbol{U}$ and $\boldsymbol{D}$. The optimization over the first subset of variables is an ℓ_1-regularized least squares problem, while the one over the second subset of variables is an ℓ_2-constrained least squares problem. In practice, there are quite a number of different algorithms available for solving this type of problems. In their paper, the authors proposed two novel fast algorithms: one named *feature-sign search algorithm* for the ℓ_1 problem and a second based on Lagrange dual for the ℓ_2 constraint. According to the reported results, both algorithms are computationally very efficient, and in particular the former outperformed the LARS and basis pursuit algorithms, well known for their superior performances in many previous works.

Remarks. Before concluding this section, we try to summarize some considerations – some methodological, while some other more practical – about a number of issues we have been confronted with while working on the preparation of this paper.

- The K-SVD algorithm proved to be very effective in practice and has the nice property that the level of sparsity of the representations is directly controlled through the ℓ_0 constraint T_0. However, one should keep in mind that the procedure adopted to solve the sparse coding step is greedy, and this may lead to some stability problems when there is a lot of correlations between the elements of the dictionary.
- In our experiments the ℓ_1-DL algorithm confirmed to be quite efficient and effective. However, despite the theoretical results on convergence to global optimum reported in the paper, we observed some slight un-expected dependence from the initialization of the dictionary. Of course these may be due to non optimal convergence criteria used in the implementation. Nonetheless, we believe this is an important point that should be kept in mind while using the framework.
- Finally, we believe that being able to be adapted easily to an on-line setting is a valuable further property of an algorithm for sparse over-complete dictionary learning in the context of medical imaging. In this respect, the ℓ_1-DL algorithm is surely better suited than K-SVD, as different works for extending the ℓ_1-based minimization to the online setting have already been proposed, and their derivation may be used.

3 Experimental Assessment

There are several specific applications where sparse and redundant representations of natural images have been shown to be very effective. The present section will demonstrate the application of such representations to the denoising of medical images, to show that, even in this well-studied domain, better results can

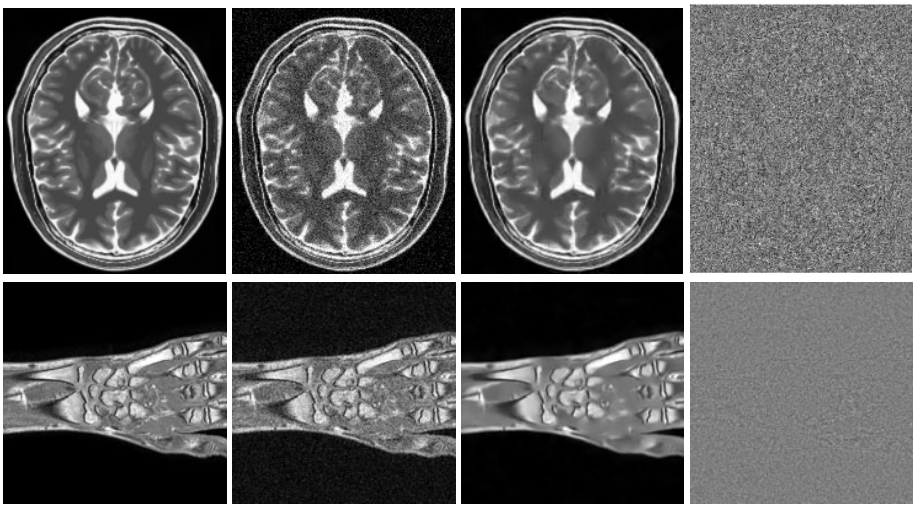


Fig. 1. Examples of denoising results on two MR images: on the top row a slice from the Brainweb dataset, on the bottom row the image of a wrist. From left to right, the first column shows the original image, the second the image with noise, the third column the denoised image (with ℓ_1-DL top and K-SVD bottom), and the fourth the difference between noisy and denoised image.

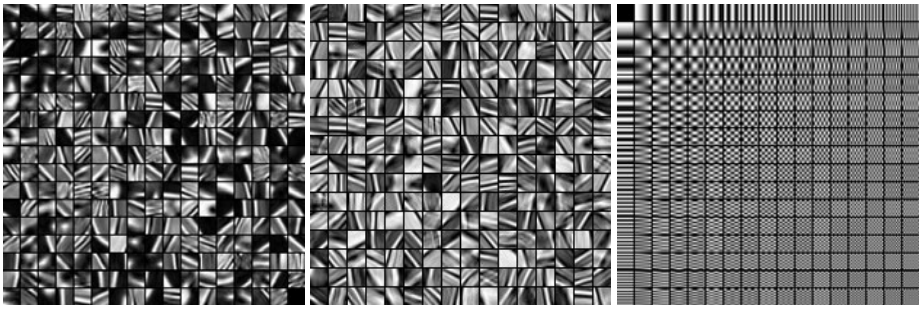
be achieved with an adaptive dictionary rather than a fixed one. However, we feel the need to emphasize that our work aims at assessing the power of the representation rather than the performance on the specific application.

Experimental Setup. The tests were carried out on MR images of three different anatomical regions: wrist, brain and kidneys (see Fig. 1 and Fig. 5). The images of the wrist have been acquired with a T1-weighted 3D Fast Gradient Echo sequence, the images of kidneys with a T2-weighted sequence, while the brain images are simulated T2-weighted, belonging to Brainweb synthetic dataset.

The experiments were based on the MATLAB code available on the page of Ron Rubinstein[1] for denoising with K-SVD. The software has been adapted to work with both K-SVD and ℓ_1-DL. All experiments were made comparing the performances of K-SVD dictionary, ℓ_1-DL dictionary (both learned from data) and a data-independent DCT dictionary. The denoising is performed iterating over all patches with a given size in the image, finding its optimal code $\boldsymbol{u}^\star$ via OMP and reconstructing the denoised patch as $\boldsymbol{D}\boldsymbol{u}^\star$, where $\boldsymbol{D}$ is the dictionary under use.

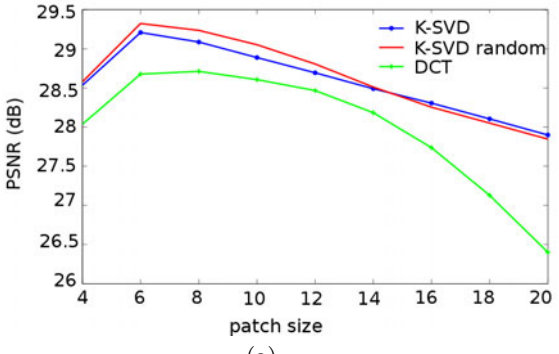
The images to be denoised were obtained adding Gaussian noise to the original ones. The experiments were made varying different parameters: the size of the patches considered, the level of sparsity of the coefficients and the level of noise to be removed. The sizes of all dictionaries are always four times the dimension of the training data, to keep a constant level of redundancy. In a first set of

[1] `http://www.cs.technion.ac.il/~ronrubin/software.html`

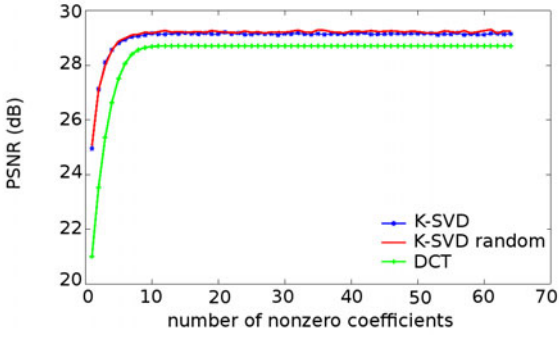


(a) (b) (c)

Fig. 2. Examples of learned dictionaries by (a) K-SVD and (b) ℓ_1-DL, compared with (c) the fixed DCT dictionary



(a)



(b)

Fig. 3. Comparison of PSNR achieved by K-SVD (initialized with DCT or randomly) and by DCT, for (a) varying sizes of the patches and (b) varying levels of sparsity

Table 1. On the left, comparison between the denoising performances of the three dictionary methods on the different images we chose. On the right, PSNR achieved by the different methods with growing level of noise.

	Kidney	Wrist	Brain
ℓ_1-DL	29.97	33.89	**29.71**
K-SVD	**30.37**	**34.14**	29.04
DCT	29.88	33.82	28.67

	$\sigma = 5$	$\sigma = 10$	$\sigma = 15$
ℓ_1-DL	**37.61**	**33.92**	**31.50**
K-SVD	35.48	32.65	30.59
DCT	34.95	32.55	30.37

experiments, the training data are 40.000 patches coming from the image to be denoised. In a second type of experiments denoising was performed with a dictionary learned from a set of noise-free images of the same anatomical region of the image to be denoised, with a total number of 60.000 training data. Denoising results were compared by measuring the peak signal-to-noise ratio (PSNR), defined as $20 \cdot \log_{10}(255\sqrt{n}/\|I - I_{den}\|)$, where n is the total number of pixels, I is the original image and I_{den} is the output of the denoising procedure, and measured in Decibel (dB).

Qualitative Assessment. A first qualitative comparison of these different dictionaries is shown in Fig. 2. Compared to DCT, the two adaptive dictionaries, both learned from the brain images, store more information about the structures

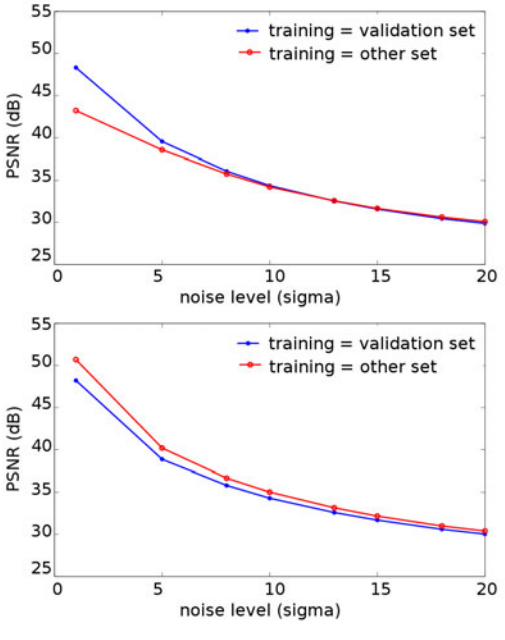


Fig. 4. Denoising performance of K-SVD (top) and ℓ_1-DL (bottom) for varying level of noise. We show the PSNR achieved by training the dictionary on the same image to be denoised, and on a different training set.

of the district we are considering. Examples of the denoising results obtained with the two dictionaries on the wrist and brain images are shown in Fig. 1. In both cases the majority of the difference image is made by noise, a sign that these methods do not disrupt the original structures of the images.

Adaptive vs Fixed. We first compared the denoising performance of an adaptive dictionary, K-SVD in this case, with respect to a non-adaptive one. In Fig. 3 is shown the variation of the PSNR of the images denoised with DCT dictionary and with two versions of the K-SVD dictionary, with different initializations. The two graphics show that adaptive dictionaries always outperformed the fixed one.

Employing an adaptive dictionary adds to the computing time an overhead due to the learning process. The parameter that most affects it in the K-SVD case is the patch size: anyway it took less than a minute to compute a dictionary with a large patch size of 18x18. Generally the patch size used is 8x8 and the number of non-zeros is 20, and on average the time taken to compute the dictionary was less than 15 seconds.

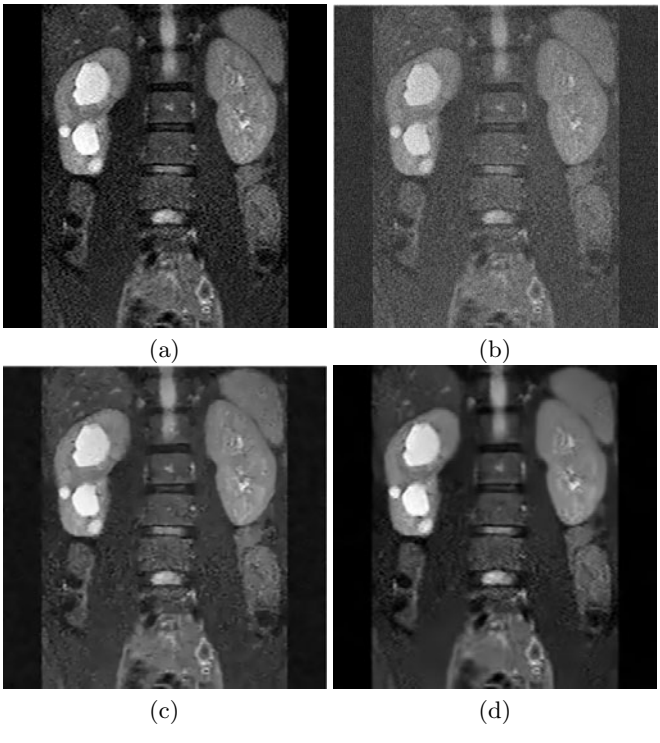


Fig. 5. Denoising on an MRI of kidneys. (a) Original image, (b) image with noise, (c) image denoised with K-SVD trained on the same noisy image, (d) image denoised after training on a different set of noise-free images. Results with ℓ_1-DL were qualitatively similar.

K-SVD vs ℓ_1-DL. The second comparison was performed between the two algorithms for dictionary learning, for a fixed patch size of 8×8 and 20 nonzero coefficients of the representation (the parameter τ in ℓ_1-DL was tuned accordingly). In Table 1 we report the results obtained on all three images for a fixed level of noise ($\sigma = 20$) and on the brain image for other three levels. The results of the DCT method are reported for the sake of completeness.

On the brain image the highest PSNR is always reached using the ℓ_1-DL dictionary, while on the other two images K-SVD achieves better results. The lowest ratio is obtained with the DCT dictionary.

A different training set. All tests above were performed using the same image for training and denoising. We tried to construct a different kind of dictionary, where the training set came from a set of medical images of the same anatomical region of the image to denoise. Experiments were performed on the images of kidneys.

The plots in Fig. 4 show the PSNR of the previous and the new version of a K-SVD dictionary (left) and of a ℓ_1-DL dictionary (both with patch size 8x8) vs the growing level of noise. In Fig. 5 we show a qualitative comparison between the reconstructions made with the two K-SVD dictionaries. In both cases at the growth of the noise level the behavior of the two dictionaries becomes more and more similar, and in fact, for $\sigma > 10$ both dictionaries could be applied to images different from the training ones without significant loss in performance. This is an important result, since it could remove the overhead due to the process of learning the dictionary on the specific image, by learning it once on a wider class of modality- and district-specific images.

4 Conclusion

This paper represents a first attempt to establish a more robust connection between medical image analysis and recent advances on sparse overcomplete coding and unsupervised dictionary learning. After an initial review of the relevant machine learning literature, we have drawn a unified presentation of the properties of different optimization approaches proposed for learning overcomplete dictionaries, and provided a more detailed description of two very promising algorithms that can be used in practice on medical images. We have concluded the paper by showing the experimental results of the selected algorithms for image denoising, which are very encouraging as they outperform the more standard technique based on DCT. We also remark that, although the experiments were run on 2D images, the extension to 3D is straightforward, a further advantage of these techniques.

References

1. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on Signal Processing [see also IEEE Transactions on Acoustics, Speech, and Signal Processing] 54(11), 4311–4322 (2006), http://dx.doi.org/10.1109/TSP.2006.881199

2. Alush, A., Greenspan, H., Goldberger, J.: Automated and interactive lesion detection and segmentation in uterine cervix images. IEEE Trans Medical Imaging 29(2) (February 2010)
3. Bishop, C.M.: Pattern Recognition and Machine Learning. Information Science and Statistics. Springer, Heidelberg (2006)
4. Combettes, P.L., Pesquet, J.C.: Proximal splitting methods in signal processing. In: Fixed-Point Algorithms for Inverse Problems in Science and Engineering. Springer, New York (2010)
5. Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionaries. IEEE Transactions on Image Processing 15, 3736–3745 (2006)
6. Elad, M., Figueiredo, M., Ma, Y.: On the role of sparse and redundant representations in image processing. Proceedings of the IEEE 98(6), 972–982 (2010)
7. Engan, K., Aase, S.O., Hakon Husoy, J.: Method of optimal directions for frame design. In: ICASSP 1999: Proceedings of IEEE International Conference on the Acoustics, Speech, and Signal Processing, pp. 2443–2446. IEEE Computer Society, Washington (1999)
8. Gonzalez, R., Woods, R.: Digital Image Processing, 3rd edn. Pearson Education, Inc., London (2008)
9. Kavukcuoglu, K., Ranzato, M., LeCun, Y.: Fast inference in sparse coding algorithms with applications to object recognition. Tech. rep., Computational and Biological Learning Lab., Courant Institute, NYU (2008)
10. Lee, H., Battle, A., Raina, R., Ng, A.: Efficient sparse coding algorithms. In: Advances in Neural Information Processing Systems 19, NIPS 2006 (2006)
11. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online learning for matrix factorization and sparse coding. Journal of Machine Learning Research 11, 19–60 (2010)
12. Mairal, J., Bach, F., Ponce, J., Sapiro, G., Zisserman, A.: Supervised dictionary learning. In: Advances in Neural Information Processing Systems 22, NIPS 2009 (2009)
13. Mallat, S.: A Wavelet Tour of Signal Processing, 3rd edn. Academic Press, New York (2009)
14. Olshausen, B., Field, D.: Sparse coding with an overcomplete basis set: A strategy employed by V1? Vision Research 37(23), 3311–3325 (1997)
15. Peyré, G.: Sparse modeling of textures. Journal of Mathematical Imaging and Vision 34(1), 17–31 (2009)
16. Raina, R., Battle, A., Lee, H., Packer, B., Ng, A.Y.: Self-taught learning: transfer learning from unlabeled data. In: Proceedings of the International Conference on Machine Learning, ICML (2007)
17. Ranzato, M., Boureau, Y., Chopra, S., LeCun, Y.: A unified energy-based framework for unsupervised learning. In: Proc. of the 11-th International Workshop on Artificial Intelligence and Statistics (AISTATS 2007), Puerto Rico (2007)
18. Rubinstein, R., Zibulevsky, M., Elad, M.: Double sparsity: Learning sparse dictionaries for sparse signal approximation. IEEE Trans. Signal Processing 58(3), 1553–1564 (2010)
19. Tropp, J., Gilbert, A.: Signal recovery from random measurements via orthogonal matching pursuit. IEEE Trans. Information Theory 53(12), 4655–4666 (2007)
20. Yang, J., Yu, K., Gong, Y., Huang, T.: Linear spatial pyramids matching using sparse coding for image classification. In: Proc. of Computer Vision and Pattern Recognition Conference, CVPR 2009 (2009)

Feature Selection for SVM-Based Vascular Anomaly Detection

Maria A. Zuluaga[1,2], Edgar J.F. Delgado Leyton[1,2], Marcela Hernández Hoyos[1], and Maciej Orkisz[2]

[1] Grupo Imagine, Grupo de Ingeniería Biomédica, Universidad de los Andes, Bogotá, Colombia
[2] CREATIS; Université de Lyon; Université Lyon 1; INSA-Lyon; CNRS UMR5220; INSERM U630; F-69621 Villeurbanne, France

Abstract. This work explores feature selection to improve the performance in the vascular anomaly detection domain. Starting from a previously defined classification framework based on Support Vector Machines (SVM), we attempt to determine features that improve classification performance and to define guidelines for feature selection. Three different strategies were used in the feature selection stage, while a Density Level Detection-SVM (DLD-SVM) was used to validate the performance of the selected features over testing data. Results show that a careful feature selection results in a good classification performance. DLD-SVM shows a poor performance when using all the features together, owing to the curse of dimensionality.

1 Introduction

Our work is motivated by computer assisted detection of coronary artery diseases (CAD) in multidetector computed tomography (MDCT) angiographic images. MDCT is increasingly used in the assessment of vascular diseases, including CAD, which remains to be the main cause of mortality worldwide [10] and is closely related to atherosclerosis. Diagnosis of the presence and severity of the CAD is critical for determining appropriate clinical management. It greatly relies on the analysis of arterial cross-sections.

Healthy arterial cross-sections usually are circular or, at least, nearly symmetric with a bar-like intensity profile in big and medium vessels, and a Gaussian-like profile for small ones. Owing to its density, homogeneity and small thickness compared to clinical image resolution, healthy coronary artery wall is hardly perceptible. Atherosclerosis affects both the arterial wall and lumen. It leads first to a thickening of the wall, then to a development of an increasingly heterogeneous plaque made up of fat, calcium, fibrous cap, etc. Such a thickened and heterogeneous wall becomes perceptible in the images (see Fig. 1). The growth of the plaque also progressively induces a narrowing of the lumen and a remodeling of its cross-sectional shape. Nevertheless, the detection and quantification of vascular lesions continue to be a tedious work for physicians who have to explore

B. Menze et al. (Eds.): MICCAI 2010 Workshop MCV, LNCS 6533, pp. 141–152, 2011.

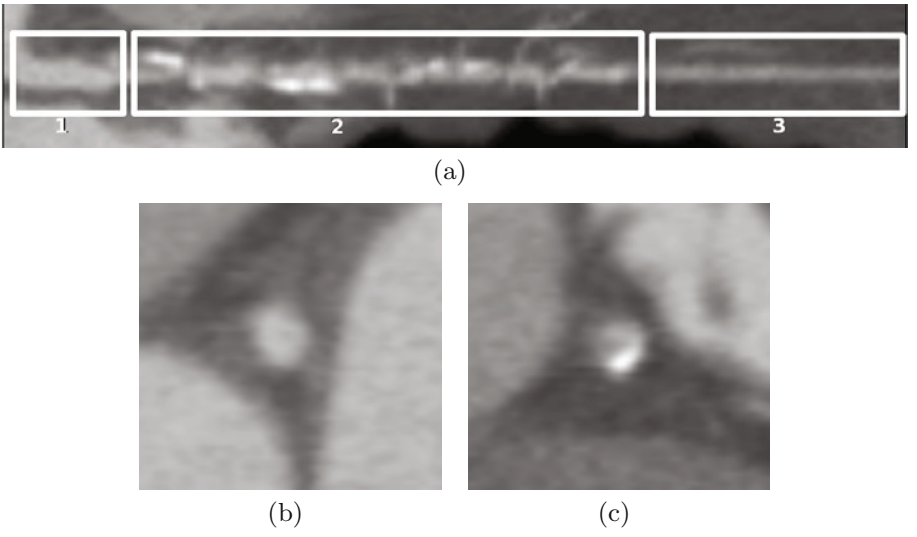


Fig. 1. Coronary artery cross-sections. (a) Curved planar reformatted view of the left anterior descending coronary artery. Rectangles 1 and 3 show healthy segments, while rectangle 2 shows a diseased segment. (b) Healthy cross-section obtained from segment 1. (c) Pathological cross-section obtained from segment 2, containing both calcified (hyperdense) and fat (hypodense) components, and a narrowed lumen.

a vast amount of data using different visualization schemes based on advanced post-processing techniques.

In a previous work [17], we proposed a method to automatically detect anomalous vascular cross-sections and attract the radiologist's attention to possible lesions, in order to reduce the time spent to analyze the image volume. We defined an intensity-based metric that aimed at differentiating normal and abnormal vascular cross-sections. The abnormality detection was defined as a Density Level Detection (DLD) problem and solved within the Support Vector Machine (SVM) scheme. Furthermore, we compared the performance of our metric with other metrics based on some classical features used in vascular enhancement and/or segmentation methods. Although our metric provided overall better results, these experiments suggested that inclusion of additional features may improve performance.

The goal of this work is twofold. First, we want to further explore feature selection to improve the performance in the vascular anomaly detection, while keeping computational times low. For this purpose, we compare several features and different feature selection strategies. Second, we want to use the obtained results to define guidelines that ease the feature selection task.

In principle, it could be thought that classification performance should increase if the number of features grows. However, the presence of inefficient features actually can degrade the classifier performance. Feature selection is not a

trivial problem. Classification performance depends on the training set size, the features and the classifier. In literature, there is a vast amount of work dedicated to feature selection [11,14,15]. Feature selection strategies can be divided into filters and wrappers. Filter methods perform a general feature selection independent of the classifier, while wrapper-type methods choose relevant features simultaneously as they conduct training/testing. A common characteristic of these methods is that they require labeling of the training data, whilst one of the advantages of the DLD-SVM method used in our work [17], is that it does not require labeled data for training. Therefore, this article follows the filter approach and attempts to determine a set of generally relevant features permitting to obtain high sensitivity and specificity in CAD detection from MDCT images, when using the DLD-SVM scheme. For this purpose, several feature selection strategies were applied on a large set of candidate features. Obviously, labeled data were necessary in this initial work at two stages. The first *labeled* subset of data was used by each feature selection strategy, in order to perform the features ranking. The best ranked features were used in the DLD-SVM scheme on *unlabeled* sets of training and validation data. Thus trained DLD-SVM was applied on testing data. As the goal was to evaluate which combination of features performs the best, these testing data were also *annotated* by experts and their labels were compared with the classification results provided by DLD-SVM.

The remaining sections are organized as follows. Section 2 describes the proposed methodology. In Section 3 experimental results are presented. Finally, discussion and conclusions are given in Sections 4 and 5 respectively.

2 Materials and Methods

The overall procedure carried out for feature selection and evaluation is summarized as follows:

1. Feature computation. MDCT coronary data sets described in Section 2.1 were used to calculate a set of features for evaluation (Section 2.2).
2. Feature selection. Various feature-selection strategies described in Section 2.3 were evaluated.
3. SVM construction. According to the feature-selection criteria of each strategy, an SVM model was constructed (Section 2.4) using training data with the selected features.
4. Performance evaluation. The classification performance over testing data was evaluated (Section 2.5).

The following sections further describe each of the above-mentioned steps.

2.1 Experimental Data

A total of 14 cardiac MDCT data sets were used: 8 from the Rotterdam Coronary Artery Algorithm Evaluation Framework [12] and the remaining 6 were acquired on a 64-row CT scanner (Brilliance 64 − Philips Healthcare, Cleveland, OH) with a standard scan protocol using the following parameters: 120

kV, 300 mAs, collimation 52×1.5 mm, rotation time 0.35 seconds and scan time 10-14 seconds. Image reconstructions were made with an in-plane pixel size of 0.37×0.37 mm^2, matrix size 512×512, slice thickness 0.9 mm, increment 0.45 mm, with an intermediate reconstruction kernel (B).

The centerlines of coronary arteries were available and used in the evaluation: 4 arteries per data set in the images from Rotterdam Coronary Artery Algorithm Evaluation Framework and 3 arteries in the remaining data sets. All features were calculated in cross-sectional planes orthogonal to these centerlines.

The obtained cross-sections were divided as follows: 1784 labeled cross-sections were used for feature selection, 3840 unlabeled cross-sections were used for DLD-SVM training and 2560 for DLD-SVM validation stage. Finally, 6400 cross-section were used for testing. The latter were annotated in order to evaluate the performance of the DLD-SVM classification with the selected set of features, but the labels were not available for the classifier.

2.2 Evaluated Features

We included in the evaluation the metric from our previous work [17] based on local features (image-intensity integrals calculated in concentric rings subdivided into sectors), as well as other metrics based on more global features commonly used in vascular enhancement and/or segmentation. Among these were: Hessian eigenvalues [3], Inertia moments [6,7], Ribbon [2], Core [4] and Ball [9]. All of these global features were computed on a multiscale basis using 8 different scales $\in [1, 1.5, 2, ..., 4.5]$ mm or sizes of the regions of interest (ROI) $\in [1 \times 1, 1.5 \times 1.5, ..., 4.5 \times 4.5]$ mm×mm. The scales/ROIs were empirically determined guaranteeing that the features had discriminative power.

Furthermore, we explored the so-named steerable features defined by Zheng *et al.* [16] in heart segmentation, which are very close to the metric proposed in [8] for vascular calcification detection. They use image derivatives (from 0 up to 2nd order) computed by convolving the data with a Gaussian kernel. As the values are sampled at points along a circular pattern, these features can also be considered as local. Owing to space limitation, we only mention the metrics and refer the reader to the original publications for detail.

2.3 Feature Selection Strategies

No selection. The first strategy consists in using all the defined features for classification using the SVMs defined in 2.4.

F-score. F-score [15] is a technique that measures the discrimination of two sets of real numbers. Given M training vectors $\boldsymbol{x}_m, m = 1, ..., M$, if the respective mean values and variances of the positive and negative sets for the ith feature are $\bar{x}^i_+, \bar{x}^i_-, (\sigma^i_+)^2$ and $(\sigma^i_-)^2$, then the F-score of the ith feature is defined as:

$$F(i) \equiv \frac{\left(\bar{x}^i_+ - \bar{x}^i\right)^2 + \left(\bar{x}^i_- - \bar{x}^i\right)^2}{(\sigma^i_+)^2 + (\sigma^i_-)^2} \tag{1}$$

where $\bar{x}^i$ is the mean value of the whole set (union of positive and negative) for the ith feature. The larger the F-score, the more likely the feature to be discriminative.

Random Forest (RF). Random forest [1] is a classification method that provides feature importance. A random forest is a classifier consisting of a collection of tree-structured classifiers, each of which is constructed by instances with randomly sampled features. Predictions are made by majority vote of the trees. The *randomForest* package of the R software[1] was used. Features were selected according to their Gini importance. Following the recommendations from Breiman [1] a large number of trees (1000) was used and the number of variables to be randomly selected from the available set of variables was the square root of the total number of features.

SVM - Recursive Feature Elimination (SVM-RFE). The SVM-RFE algorithm [5] returns the ranking of the features of a classification problem by training a SVM with a linear kernel and removing the feature with smallest ranking criterion at each step. An implementation based on the LIBSVM[2] interface for R software was used. A coarse grid search was performed over parameter $C \in \{0.001, 0.01, ..., 100\}$ showing no significant variation in the results. C=10 was kept for the experimentations.

2.4 Support Vector Machine Classification

As mentioned before, we use a Support Vector Machine (SVM) scheme that does not require labeled training data, which represents an advantage since labeled data are not easy to obtain in the medical domain. This approach was defined by Steinwart *et al.* [13] who formulated the anomaly detection as a DLD classification problem and solved it by use of SVM. In the DLD formulation, anomalies are defined as points with a low probability density value, *i.e.* belonging to the set $\{h \leq \rho\}$ of points below a threshold level $\rho > 0$. The complement $\{h > \rho\}$ is called the normal set. Here h is an unknown density $h = dQ/d\mu$ of an unknown probability measure Q that describes the data, with respect to a known reference measure μ, both defined and i.i.d. on a space $\mathbf{X}$. The goal of the DLD problem is to find an estimate of the anomalous set, which is done indirectly, by estimating the normal set. Interpreting DLD as a classification problem, it is possible to construct an SVM that solves it. According to [13], for a DLD problem with a training set $\mathcal{T} = \{T, T'\}$ (where T and T' respectively are samples from Q and μ), a regularization parameter $\lambda > 0$, and $\rho > 0$, the SVM chooses a decision function $f_{\mathcal{T},\mu,\lambda}$ that minimizes in the Hilbert space $\mathcal{H} \times \mathbb{R}$ the following criterion:

$$\lambda \parallel f \parallel^2_{\mathcal{H}} + \frac{1}{(1+\rho)|T|} \sum_{\boldsymbol{x}_m \in T} l(1, f(\boldsymbol{x}_m)) + \frac{\rho}{1+\rho} \mathbb{E}_{X \sim \mu} \left(l(-1, f(\boldsymbol{x})) \right), \quad (2)$$

[1] http://www.r-project.org/

[2] http://www.csie.ntu.edu.tw/~cjlin/libsvm

where l is the hinge loss function and an approximation of the expectation $\mathbb{E}_{X\sim\mu}\left(l(-1, f(\boldsymbol{x}))\right)$ is calculated using the set T'. The reproducing kernel Hilbert space k is built using Gaussian radial base functions with variance σ^2. Two parameters need to be determined: λ and σ^2. This is done by a grid-search procedure performed over the two parameters. However, since there is no labeling of the training data, the grid search aims at minimizing an empirical risk function. The empirical risk of f with respect to a pair $\{T, T'\}$ is defined as:

$$\mathcal{R}_{T,T'}(f) = \frac{1}{(1+\rho)|T|}\sum_{\boldsymbol{x}\in T} l\left(1, \text{sign}(f(\boldsymbol{x}))\right) + \frac{\rho}{(1+\rho)|T'|}\sum_{\boldsymbol{x}\in T'} l\left(-1, \text{sign}(f(\boldsymbol{x}))\right), \tag{3}$$

where $l(\cdot)$ denotes the loss function from the standard SVM classification.

2.5 Performance Evaluation

The feature selection performance was evaluated as follows: first the DLD-SVM method was used to train models with the best ranked features from each strategy, then these models were applied on the testing sets, finally thus obtained classification results were used to compute the following performance measure:

$$BER = \frac{1}{2}\left(\frac{|\text{erroneous } \boldsymbol{x}_+|}{|\boldsymbol{x}_+|} + \frac{|\text{erroneous } \boldsymbol{x}_-|}{|\boldsymbol{x}_-|}\right), \tag{4}$$

called balanced error rate (BER). BER is the average of the error rates of the positive $\{\boldsymbol{x}_+\}$ and the negative $\{\boldsymbol{x}_-\}$ classes, and is related to the area under the ROC curve (AUC). The ROC curve is obtained by varying a threshold on the discriminant values (outputs) of the classifier. The curve represents the fraction of true positive as a function of the fraction of false negative. For classifiers with binary outputs, we have: $AUC = 1 - BER$. Additionally, computational times were measured since this is a critical issue in clinical practice.

3 Results

Before the actual comparison of different candidate features and feature selection strategies we had to reduce the dimensionality of the steerable features [16] used in our evaluation. Indeed, these features represent a huge amount of information, and depend on the location, at which they are calculated. For example, the gradient might be a good feature at points close to a calcification but irrelevant elsewhere. Our initial implementation contained a total of 1041 sampling points and 20 different measures (see Table 1). This represented a total of 20820 values. In order to reduce the dimensionality of the steerable features and to evaluate the discriminative power of each measure, a separate evaluation of this criterion was done using the F-score. Mean, median, maximal and minimal values, as well as variance of the F-score of the 20 measures computed at sample points were calculated. Table 1 summarizes the obtained values, but omits the minimum since there was no significant difference between measures.

Table 1. F-score summary for steerable features. Features of order zero (I), one (I_x, I_y and $|\nabla I|$) and two (I_{xx}, I_{yy}, I_{xy} and $||\boldsymbol{H}||_F$) are presented.

Measure	Mean $\times 10^{-2}$	Median $\times 10^{-2}$	Max. $\times 10^{-2}$	σ^2 $\times 10^{-2}$	Measure	Mean $\times 10^{-2}$	Median $\times 10^{-2}$	Max. $\times 10^{-2}$	σ^2 $\times 10^{-2}$
I	6.6	6.2	16.4	0.2	$\|\nabla I\|^3$	4.0	2.2	21.7	0.2
$\sqrt{I}$	6.5	6.2	16.7	0.2	$\log \|\nabla I\|$	4.9	3.6	19.5	0.2
I^2	6.5	6.2	15.7	0.2	I_{xx}	7.2	4.9	28.3	0.5
I^3	6.2	6.0	16.3	0.2	I_{yy}	8.0	5.6	29.1	0.6
$\log I$	6.4	6.0	16.8	0.2	I_{xy}	8.5	5.1	38.5	0.8
I_x	8.5	5.1	37.7	0.8	$\|\|\boldsymbol{H}\|\|_F$	3.8	1.2	29.3	0.4
I_y	8.0	4.6	40.8	0.8	$\sqrt{\|\|\boldsymbol{H}\|\|_F}$	3.7	1.1	27.4	0.4
$\|\nabla I\|$	5.9	4.2	26.0	0.4	$\|\|\boldsymbol{H}\|\|_F^2$	3.4	1.5	26.8	0.3
$\sqrt{\|\nabla I\|}$	5.8	4.2	23.7	0.3	$\|\|\boldsymbol{H}\|\|_F^3$	2.7	1.5	19.3	0.2
$\|\nabla I\|^2$	5.2	3.0	25.8	0.3	$\log \|\|\boldsymbol{H}\|\|_F$	3.4	1.1	23.3	0.3

Results from Table 1 show that there are several measures having a similar discriminative power. In particular, for a given derivative order, the modification of the value by an operator (*i.e.* power, logarithm) does not increase the discriminative power. For this reason, we excluded these redundant measures. Eight measures were kept from steerable features: I, I_x, I_y, $|\nabla I|$, I_{xx}, I_{yy}, I_{xy} and $||\boldsymbol{H}||_F$ (where $||\boldsymbol{H}||_F$ is the Frobenius norm of the Hessian), which represented a reduction in the number of values to 8328. An initial feature evaluation with the RF and SVM-RFE methods including the reduced set of steerable features was highly time consuming (more than 8 hours) and provided very poor results. This was attributed to the curse of dimensionality problem since the number of features was significantly higher, in magnitude, than the amount of data (1784) devoted to feature selection. Therefore, it was necessary to perform an additional feature size reduction.

The adopted strategy consisted in sub-sampling the spatial locations at which the different measures were evaluated. An incremental sampling rate $\in [2, 3, 4, 10, 15, 21]$ was applied, while evaluating the F-score on each new subset to guarantee that the discriminative power was not lost. Results showed that, even for a high sampling rate, the discriminative power of these features was not significantly altered (Fig. 2(a)). This allowed an additional reduction to 400 values (50 sampling points and 8 measures).

The average F-score response for the concentric rings metric and global features is summarized in Figs. 2 (b)-(c). To simplify information, ring metric sectors are averaged to obtain a score on a ring basis. Similarly, the F-scores of different λ values are averaged at each scale, for the features based on Hessian eigenvalues and inertia moments. These figures demonstrate that the discriminative power of these features depends on the scale/ROI or ring radius respectively. Nevertheless, all scales/ROIs and rings were kept in our experimentation.

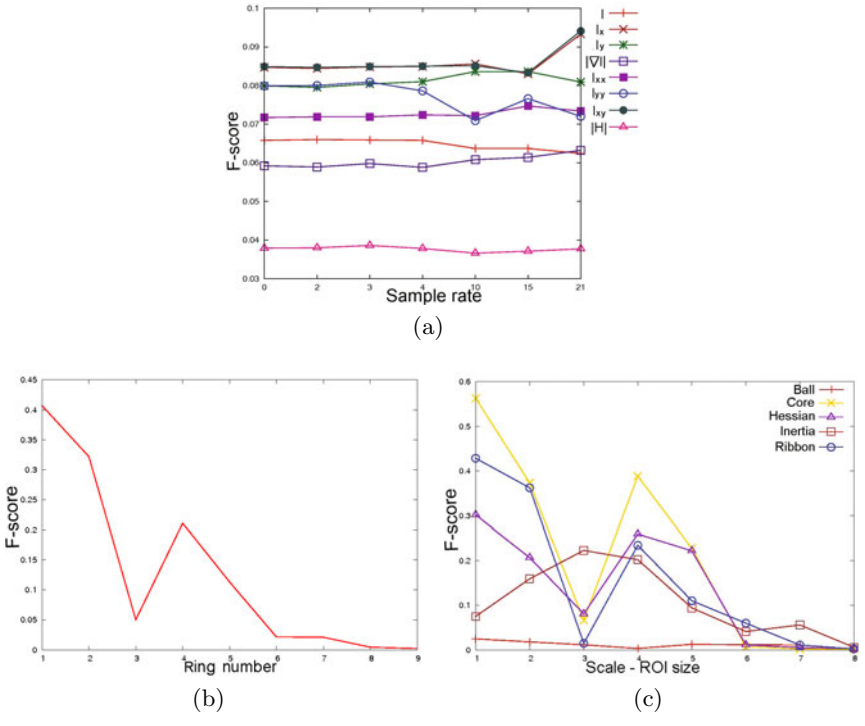


Fig. 2. Average F-scores. (a) For 8 steerable features: I, I_x, I_y, $|\nabla I|$, I_{xx}, I_{yy}, I_{xy} and $||\boldsymbol{H}||_F$, as a function of the sampling rate of spatial locations. (b) For concentric rings metric (abscissa = ring number). (c) For global measures as a function of scale/ROI.

Table 2. Top-10 feature ranking according to F-score, RF and SVM-RFE strategies

Strategy	1	2	3	4	5	6	7	8	9	10
F-score	Core	Ribbon	Hessian	Rings	Inertia	I_x	I_{xy}	I_y	I_{yy}	I_{xx}
RF	Core	Ribbon	Rings	I_x	I_y	I_{yy}	I_{xx}	I_{xy}	Inertia	Hessian
SVM-RFE	Core	Ribbon	Rings	Hessian	I_y	I_x	$\|\nabla I\|$	I_{xx}	I_{yy}	I_{xy}

3.1 Feature Selection

Since each feature contains a considerable amount of measures (*i.e.* each steerable feature contains 50 values corresponding to 50 spatial locations), the final ranking criterion was defined by averaging the ranks of every measure associated to a given feature. More precisely, the ranks of the global features were averaged over scales/ROIs, while the ranks of the local features were averaged over spatial locations. In all selection strategies, the worst metric was Ball, followed by image intensity I. The 10 best ranked features according to each selection strategy are

shown in Table 2. It can be noted that these top-10 features were exactly the same for F-score and RF strategies, while SVM-RFE preferred the gradient magnitude ∇I to the Inertia moments. Therefore, the performance evaluation in the DLD-SVM scheme would provide identical results for F-score and RF selection strategies, if the top-10 features were kept in both cases. However, the actual ranks of the features differed from one strategy to another and the F-scores of the steerable features were significantly smaller than those of the global features and of the concentric rings. For these reasons, in the case of F-score based selection, we included in the final evaluation only the features with the score greater than 0.1, which rejected all steerable features and Ball.

3.2 Performance of DLD-SVM with the Selected Features

The performance on testing data, using the best set of features selected by each strategy, is summarized in Table 3. Results obtained using F-score and RF are pretty close, while the ones obtained with SVM-RFE present a higher BER. Results in terms of BER are also presented for the DLD-SVM framework using only the concentric ring metric. It can be seen that the inclusion of new features improves the performance with respect to our previous work. Please note that different features involve different numbers of measures, which partly explains the differences of computational time.

Table 3. Performance of each feature set. BER values are presented as a percentage, while computational time is expressed in minutes. The last line reports the result of our previous work using the concentric rings metric alone.

Strategy	BER	Computational time
No strategy	40.1	1625
F-score	15.24	80
RF	17.21	21
SVM-RFE	27.67	255
DLD-SVM with concentric rings	23.38	75

4 Discussion

Feature selection using three different methods showed several coincidences. All strategies agree in the worst placed features: Ball measure and image intensity. Nevertheless, the Ball measure might be helpful in detecting aneurysms. As the experimental data did not contain this type of anomalies, it is not surprising that this measure did not provide valuable information. As for image intensity, it can vary significantly from one data set to another or even in the same image (*e.g.* lumen intensity varies significantly from proximality to distality). This high variability can explain the bad rank. Regarding highest ranked features, Ribbon and Core metrics were always among the best. These two measures are

similar to the concentric rings metric, which demonstrated a good classification performance in our previous work.

Despite their frequent use in vascular image enhancement and segmentation, Hessian eigenvalues and inertia moments show a relatively low ranking, the latter were even rejected by SVM-RFE. One possible explanation is that both metrics usually aim at describing the shape of the lumen, which does not imply they can detect changes in the vessel wall appearance. Additionally, Hessian eigenvalues are sensitive to noise, while inertia moments integrate voxels intensities regardless of whether or not they belong to lumen, wall or background. Consequently, the inertia moments possibly are not sensitive enough to local changes in lumen and wall, while they detect global changes in the surroundings. Actually, the measures that perform the best, use integration to reduce noise sensitivity, but this integration is performed locally, *e.g.* within rings and sectors. Another piece of the explanation is that both Hessian eigenvalues and inertia moments usually are computed in 3D, which also provides information on local elongation, but is more time-consuming. For fair comparison with other 2D features, we only could exploit their capability of describing the cross-sectional shape.

Results on testing data demonstrated that F-score and RF feature selection strategies do improve performance. The similarity in their BER response can be explained by the close relationship between the feature sets selected by both methods. Additional features selected by RF did not improve and even slightly degraded the classification. For SVM-RFE, further investigation is needed to understand why the selected features perform worse than the ones obtained from the other two methods. Finally, the poor performance of the DLD-SVM when using all the features can be justified by the curse of dimensionality.

5 Conclusions

In this paper, we have evaluated a wide set of features for the purpose of vascular anomaly detection. Three different strategies were used to determine the best set of features. More than determining a particular set of features, our research has allowed us to create some guidelines that may be followed when selecting features for this type of problem. These can be summarized as follows:

1. Integral-based features show a good performance provided that they are semi-local, such as Core, Ribbon and Rings.
2. Proper scale definition or ROI selection is crucial. Typically, the maximum scale/ROI should be of the same order as the vessel diameter. Larger scales should be avoided. They tend to include spurious information without discriminant power that affects the classifier performance.
3. Fine sampling of local features is not mandatory. We demonstrated that their coarse sampling has the same discriminative power as a fine grid.
4. When using derivatives, rather than the derivative magnitudes prefer their components, as they showed more relevance in the classification task.
5. Avoid the direct use of the image intensity, since it has a high variability that reduces its discriminative power.

It is important to remark that in feature selection the discriminative power of a feature is not the only factor to be taken into account. Mutual information among features is important. Therefore the selection of a single highly discriminant feature, *i.e.* Cores, can have a worse performance than a set of not so discriminant features. As an example, although in previous works the concentric rings provided overall good results, these experiments suggested that inclusion of additional features improve performance. Another contribution of the paper is that we empirically demonstrated that careful feature selection results in a good classification performance. In particular, DLD-SVM shows a poor performance when using all the features, due to the curse of dimensionality. As for SVM-RFE feature selection strategy, further investigation is needed to understand why the selected features perform worse than the ones obtained from the other two methods.

While our work required the use of a large number of labeled data, for the purpose of evaluation of different features combinations, the actual anomaly detection method does not require labeled data. The interested reader only needs to implement the DLD-SVM method and the computation of the recommended features that are used as inputs of the classifier. Future work can focus in extending the evaluation by inclusion of other arteries, inclusion of additional features and a deeper evaluation on the effect of the features number in classification.

Acknowledgments. This work has been supported by the ECOS-Nord Committee, project C07M04, by the Région Rhône-Alpes (France) via the Simed project of the ISLE research cluster, and by the project CIFI-Uniandes No. 54. M.A. Zuluaga's PhD project is supported by a Colciencias grant. We are very grateful to Dr. Don Hush for his kind advices.

References

1. Breiman, L.: Random forests. Mach. Learning 45(1), 5–32 (2001)
2. Florin, C., Paragios, N., Williams, J.: Particle Filters, a Quasi-Monte Carlo solution for segmentation of coronaries. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 246–253. Springer, Heidelberg (2005)
3. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale vessel enhancement filtering. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 130–137. Springer, Heidelberg (1998)
4. Fridman, Y., Pizer, S.M., Aylward, S., Bullitt, E.: Segmenting 3D Branching Tubular Structures Using Cores. In: Ellis, R.E., Peters, T.M. (eds.) MICCAI 2003. LNCS, vol. 2879, pp. 570–577. Springer, Heidelberg (2003)
5. Guyon, I., Weston, J., Barnhil, S., Vapnik, V.: Gene Selection for Cancer Classification using Support Vector Machines. Mach. Learning 46(1-3), 389–422 (2002)
6. Hernández Hoyos, M., Orkisz, M., Douek, P.C., Magnin, I.E.: Assessment of carotid artery stenoses in 3D contrast-enhanced magnetic resonance angiography, based on improved generation of the centerline. Mach. Graphics and Vision 14(4), 349–378 (2005)

7. Hernández Hoyos, M., Serfaty, J.M., Maghiar, A., Mansard, C., Orkisz, M., Magnin, I.E., Douek, P.C.: Evaluation of semi-automatic arterial stenosis quantification. Int. J. of Computer Assisted Radiol. and Surg. 1(3), 167–175 (2006)
8. Išgum, I., Rutten, A., Prokop, M., van Ginneken, B.: Detection of coronary calcifications from computed tomography scans for automated risk assessment of coronary artery disease. Med. Phys. 34(4), 1450–1461 (2007)
9. Nain, D., Yezzi, A., Turk, G.: Vessel Segmentation Using a Shape Driven Flow. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3216, pp. 51–59. Springer, Heidelberg (2004)
10. World Health Organization: The top ten causes of death - Fact sheet N310 (October 2008)
11. Peng, H., Long, F., Ding, C.: Feature Selection Based on Mutual information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. IEEE Trans. Pattern Anal. and Mach. Intell. 27(8), 1226–1238 (2005)
12. Schaap, M., Metz, C., van Walsum, T., van der Giessen, A.G., Weustink, A.C., Mollet, N.R., Bauer, C., Bogunovic, H., Castro, C., Deng, X., Dikici, E., O'Donnell, T., Frenay, M., Friman, O., Hernández Hoyos, M., Kitslaar, P.H., Krissian, K., Kühnel, C., Luengo-Oroz, M.A., Orkisz, M., Smedby, Ö., Styner, M., Szymczak, A., Tek, H., Wang, C., Warfield, S.K., Zambal, S., Zhang, Y., Krestin, G.P., Niessen, W.J.: Standardized Evaluation Methodology and Reference Database for Evaluating Coronary Artery Centerline Extraction Algorithms. Med. Image Analysis 13(5), 701–714 (2009)
13. Steinwart, I., Hush, D., Scovel, C.: A Classification Framework for Anomaly Detection. J. Mach. Learning Research 6, 211–232 (2005)
14. Wróblewska, A., Boniński, P., Przelaskowski, A., Kazubek, M.: Segmentation and feature extraction for reliable classification of microcalcifications in digital mammograms. Opto-electronics Review 11(3), 227–235 (2003)
15. Yang, X., Jia, L., Qingmao, H., Zhijun, C., Xiaohua, D, Pheng Ann, H.: F-score Feature Selection Method May Improve Texture-based Liver Segmentation Strategies. In: Proceedings of 2008 IEEE Int. Symp. on IT in Med. and Education, pp. 697–702 (2008)
16. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Fast automatic heart chamber segmentation from 3D CT data using marginal space learning and steerable features. In: Int. Conf on Computer Vision, pp. 1–8 (2007)
17. Zuluaga, M.A., Magnin, I.E., Hernández Hoyos, M., Delgado Leyton, E.J.F., Lozano, F., Orkisz, M.: Automatic detection of abnormal vascular cross-sections based on Density Level Detection and Support Vector Machines. Int. J. Comput. Assisted Radiol. Surg. (in-press, 2010), doi:10.1007/s11548-010-0494-8

Multiple Classifier Systems in Texton-Based Approach for the Classification of CT Images of Lung

Mehrdad J. Gangeh[1], Lauge Sørensen[2], Saher B. Shaker[3], Mohamed S. Kamel[1], and Marleen de Bruijne[2,4]

[1] Department of Electrical and Computer Engineering, University of Waterloo, Canada
{mgangeh,mkamel}@pami.uwaterloo.ca
[2] Department of Computer Science, University of Copenhagen, Denmark
{lauges,marleen}@diku.dk
[3] Department of Respiratory Medicine, Gentofte University Hospital, Hellerup, Denmark
[4] Biomedical Imaging Group Rotterdam, Erasmus MC, The Netherlands

Abstract. In this paper, we propose using texton signatures based on raw pixel representation along with a parallel multiple classifier system for the classification of emphysema in computed tomography images of the lung. The multiple classifier system is composed of support vector machines on the texton signatures as base classifiers and combines their decisions using product rule. The proposed approach is tested on 168 annotated regions of interest consisting of normal tissue, centrilobular emphysema, and paraseptal emphysema. Texton-based approach in texture classification mainly has two parameters, i.e., texton size and k value in k-means. Our results show that while aggregation of single decisions by SVMs over various k values using multiple classifier systems helps to improve the results compared to single SVMs, combining over different texton sizes is not beneficial. The performance of the proposed system, with an accuracy of 95%, is similar to a recently proposed approach based on local binary patterns, which performs almost the best among other approaches in the literature.

1 Introduction

Texture-based pixel classification in computed tomography (CT) images of the lung is an emerging and promising tool for quantitative analysis of lung diseases such as emphysema, one of the main components of chronic obstructive lung disease (COPD). Emphysema, which is characterized by the loss of lung tissue, is visible in CT images as textural patterns comprising low intensity blobs or low attenuation areas of varying sizes (refer to Fig. 1 for some examples). Supervised classification using texture features allows for taking the emphysema texture into account. This was introduced in [1], and since then, various features have been used for the classification of emphysema and other abnormalities in lung CT images, including moments of filter response histograms from filter banks based on Gaussian derivatives [2], measures on gray-level co-occurrence matrices (GLCM), measures on gray-level run-length matrices (GLRLM), and moments of the attenuation histogram [1, 3, 4].

B. Menze et al. (Eds.): MICCAI 2010 Workshop MCV, LNCS 6533, pp. 153–163, 2011.

It has been shown, recently, that small-sized local operators like local binary patterns (LBP) [5] and patch representation of small local neighborhood in texton-based approaches [6] yield excellent texture classification performance on standard texture databases. Small-sized local operators are especially desirable in situations where the region of interest (ROI) is rather small, which is often the case in texture analysis in medical imaging, where pathology can be localized in small areas. This is because of two reasons: first, convolution with large support filter banks suffers from boundary effects; second, more patches can be extracted using small-sized local operators that makes the estimation of image statistics more reliable [6]. It is also shown in [7] that using support vector machines (SVMs) instead of k nearest neighbor (k-NN) classifier, which is used in [6, 8] can improve the performance of texton-based approaches even further.

In our previous work [9], we proposed to use small patch representation in texton-based approaches for the classification of emphysema in CT images of the lung. This approach mainly consists of two *learning* steps: first an unsupervised step using k-means to construct a codebook of textons. Second, a supervised step in which the model is learned by obtaining a histogram of textons to represent each region of interest (ROI). There are two main parameters in these two steps, k in k-means and texton size (TS), i.e., the size of patches extracted from the ROIs.

In general, the optimal parameters can vary regionally within the lung and from patient to patient, depending on the intrinsic scale and complexity of the texture patterns. Hence, it is not known *a priori* which texton size or k value in k-means yields the best performance [10]. Hence, one possibility is to represent the ROIs using various k and TS parameters that yield several feature subsets. There are several approaches in the literature to incorporate the information from all these feature subsets into a single decision system. These are including combined feature space (CFS), in which the feature subsets are fused to construct a single feature space and then submit it to a single classifier [11]; multiple classifier systems (MCS), where each feature subset is submitted to a so called base classifier (BC) and the decision by these BCs are combined into a single decision [11]; and multiple kernel learning (MKL), a system of multiple support vector machines (SVMs), each of which with its own kernel [12, 13]. In MKL, the weights used for combining the decisions of the SVMs are optimized within the SVM optimization which leads to a quadratic optimization problem with quadratic constraint.

In this paper we propose to use multiple classifier systems with SVM as base classifier to aggregate the decisions made by the base classifiers using features (here histograms) at multiple texton sizes or multiple k values. However, all SVMs use the same kernel, a radial basis function (RBF) kernel, and the decisions by these SVMs are combined using a fixed rule such as product rule. Our results show that the performance of the classification system using multiple classifier systems produces better results than single base classifiers and provides a means for making use of the information at various parameters of the approach. It also yields similar to or better results than the current approaches in the literature for the same application such as local binary patterns (LBPs) or filter bank approaches.

2 Texton-Based Approach

In this section, we first briefly explain texton-based approach in texture classification. Then we present multiple classifier systems at multiple texton-based features as a means to aggregate feature subsets obtained at various texton sizes or k values.

2.1 Texton-Based Texture Classification

The basic idea of textons was first introduced by Julesz as the elements of texture perception [14]. However, it took sometime before this idea could be developed into a texture classification system as proposed in [15]. This technique was further improved by Cula and Dana [16] and also Varma and Zisserman [6, 8] that yielded higher performance on standard texture databases.

There are three main representations associated with the texton-based approach in the literature, i.e., filter banks [8, 15, 16], raw pixel representation [6], and Markov random field (MRF) representation, where the central pixel in a neighborhood is modeled using the neighboring pixels [6]. However, irrespective of the representation used to describe local image information, the texton-based approach consists of learning and classification stages [6]. The learning stage, in turn, is divided into three steps: 1) construction of a codebook of textons using an unsupervised (clustering) algorithm such as k-means; 2) learning texton histograms from the training set; and 3) training a classifier such as SVM using texton histograms obtained in step two. In the classification stage, the class of a test image is determined by submission of the histogram of textons in the test image to the classifier trained in the learning stage.

To construct the texton codebook, small-sized local patches are randomly extracted from each image[1] in the training set. These small patches are then converted to the appropriate representation such as filter banks or raw pixels. Eventually, they are aggregated over all images in a class and clustered using a clustering algorithm such as k-means. The cluster centers obtained form a dictionary of textons to represent the class of textures. It will be used as the codebook of textons in the next step. The size of the dictionary depends on the number of cluster centers, e.g., k in k-means algorithm as well as the number of classes. For example, for a three-class problem with k of 30, $3 \times 30 = 90$ textons are generated in the codebook. Fig. 1 displays sample images of lung CT ROIs used in this paper as well as a codebook of 90 textons computed over all ROIs using the texton size of 9×9 pixels and $k = 30$.

The second step in the learning stage is supervised, in which a histogram of textons is found for each image in the training set as a model (feature set) to represent this image. To find this histogram, small patches of the same size as in the unsupervised step are extracted by sliding a window over each training image in a class. These patches are then converted to the appropriate representation as used in the previous step. Finally, a histogram of textons is computed for the image by comparing each and every patch representation in that image with all textons in the dictionary using a similarity measure to find the closest match and updating the corresponding histogram bin based on the closest match found. The histograms are normalized and used as the feature sets for the images in the training set and employed for training a classifier

[1] In this paper image and region of interest (ROI) of the lung are used interchangeably.

such as a support vector machine (SVM) as the third step of the learning. Left and middle diagrams in Fig. 2 illustrate the construction of the codebook and learning the model in a texton-based classification system using raw pixel representation.

In the classification stage, to classify a test image, the same steps as in the learning stage are followed to find the features for the test image. This includes extraction of small patches from each test image in a class, converting the patches to the appropriate representation, finding the closest match to these patches from the dictionary, and computing the normalized histogram of obtained closest textons to define a feature vector for the image. The trained classifier in the learning stage is subsequently used to find the class of the test image. In SVM, a RBF kernel as given in (1) is used as it is recommended as the first kernel choice in [17]. In (1), γ is the kernel width and $\mathbf{x}_i$ and $\mathbf{x}_j$ are two sample patterns.

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2}. \tag{1}$$

2.2 Multiple Classifier Systems

The learning stage in texton-based approach generates an n-dimensional vector $\mathbf{h}^{(i)} = [h_1, \ldots, h_n] \in \mathbb{R}^n, i = 1, \ldots, m$ for each ROI, where n is the number of bins in the histogram of textons and m is the total number of texton-sizes or k values for which the model is learned. Each $\mathbf{h}^{(i)}$ is considered as a feature subset obtained at a specific texton size or k value and they can be composed into a single feature space $\mathbf{h} = [\mathbf{h}^{(1)}, \ldots, \mathbf{h}^{(m)}]$, which is called *distinct pattern representation* (DPR) [18].

We propose here to submit this DPR to an ensemble of classifiers [11]:

$$\Gamma = \{D_1, \ldots, D_m\}, \qquad \Gamma: \mathbb{R}^{n \times m} \rightarrow \Omega^m \tag{2}$$

where, Γ is the ensemble with $D_i: \mathbb{R}^n \rightarrow \Omega, i = 1, \ldots, m$, as base classifier (BC) trained on each feature subset $\mathbf{h}^{(i)} \in \mathbb{R}^n, i = 1, \ldots, m$ and $\Omega = \{\omega_1, \ldots, \omega_c\}$ is the set of class labels.

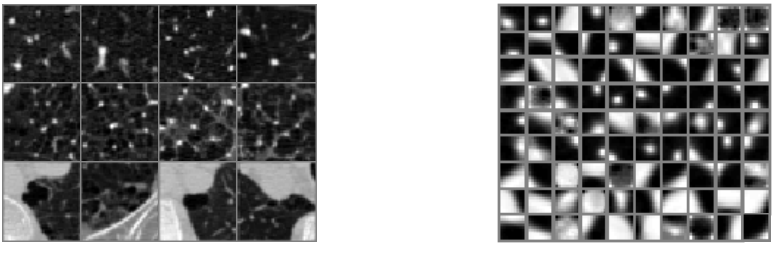


Fig. 1. Sample ROIs of size 50 × 50 pixels (*left*) in three classes, i.e., normal lung (*top left row*), CLE (*middle left row*), and PSE (*bottom left row*). The constructed codebook using texton sizes of 9 × 9 pixels and $k = 30$ in k-means (*right*).

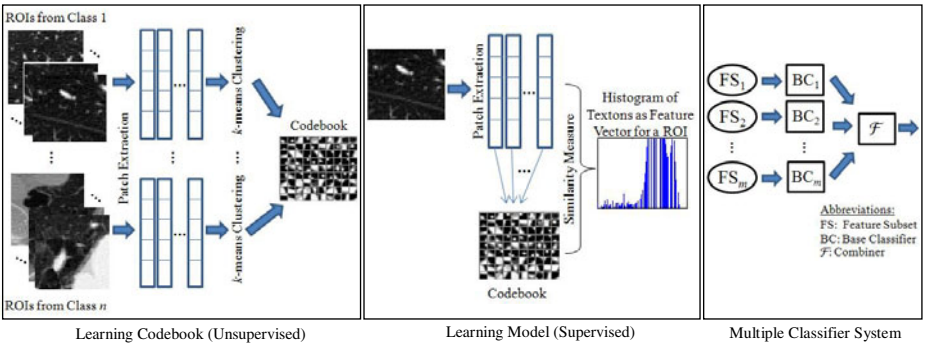


Fig. 2. The illustration of different stages of the proposed system using multiple classifier systems and texton signatures: the generation of texton codebooks using *k*-means clustering (*left*), the generation of features by computing the texton histograms (*middle*), and parallel multiple classifier system to aggregate the decisions by single base classifiers (*right*)

The decisions made by these BCs are subsequently fused by the aggregation function $\mathcal{F}$ to yield a single decision on the class of the pattern submitted for classification such that $\mathcal{F}: \Omega^m \rightarrow \Omega$.

There are three main structures of multiple classifier systems (MCS), i.e., stacked MCS with the same feature space for all BCs; parallel MCS with a distinct feature space for each BC; and sequential MCS in which the output of each BC is given to the next one. Here since the feature subset $\mathbf{h}^{(i)}$ given to each BC is different, parallel MCS is a natural choice. The right diagram in Fig. 2 illustrates the structure of the proposed multiple classifier system.

3 Experimental Setup

Data Preparation. Emphysema is often classified into various subtypes based on morphology [19]. In this work, we focus on the two subtypes related to smoking, namely, centrilobular emphysema (CLE), defined as multiple small low-attenuation areas and paraseptal emphysema (PSE), defined as multiple low-attenuation areas in a single layer along the pleura often surrounded by interlobular septa that is visible as thin white walls. The data used for the experiments is the same as in [9, 20, 21] and comprises 168 ROIs, of size 50 × 50 pixels, representing the following three classes: normal tissue (NT) (59 ROIs), CLE (50 ROIs), and PSE (59 ROIs). The ROIs are extracted from 75 thin-slice pulmonary CT images of 25 different subjects where the leading pattern was obtained as the consensus visual assessment by two experienced readers. The NT ROIs are from healthy non-smokers while the emphysema ROIs are from smokers. CT was performed using GE equipment (LightSpeed QX/i; GE Medical Systems, Milwaukee, WI, USA) with four detector rows, using the following parameters: in-plane resolution 0.78 × 0.78 mm, 1.25 mm slice thickness, tube voltage 140 kV, and tube current 200 mAs. The slices were reconstructed using a high spatial resolution (bone) algorithm.

Computation of Texton-Based Features. The codebook of textons is constructed by extracting 500 random patches from each ROI in the training set. Patch sizes of 3×3 to 9×9 pixels are used in the experiments. Raw pixel representation is used. Since in CT images, the mean of the intensity in the images indicate a physical property of the tissue, the mean of the ROIs are not removed. The patches extracted from different ROIs of each class are given to k-means clustering algorithm to find the codebook. Five different values of k, i.e., $k = 10$ to $k = 50$ are tested in the experiments. After construction of texton codebook, texton frequency histograms of texture images are computed to find the model. In this stage, small overlapping patches with the same size as what was used in the clustering stage are extracted from top left to bottom right of each ROI. As in the clustering stage, no filter bank is used and raw pixel representation is considered. The Euclidean distance between the resulting textons (collection of small patches) and textons in the codebook is computed in order to identify the most similar texton in the codebook and the corresponding histogram of textons is updated. Normalized histograms are used as the feature subsets $\mathbf{h}^{(i)}$.

Classifier and Evaluation. The evaluation of the classification system is performed using leave-one-subjectp-out. This means that all the ROIs of one subject (patient) are used as the test set and the remaining ROIs as the training set. A parallel multiple classifier system with SVM as base classifier (BC) is used. It is shown in [9] that SVM performs better than k-NN in the classification of emphysema and hence SVM is used as the BC in our experiments. Product combiner is selected as the aggregation function $\mathcal{F}$as our preliminary experiments show that it performs almost the best among other combiners including majority voting, mean, and max combiners. The crucial issue in using SVM is finding a suitable kernel and the optimum trade-off parameter C. RBF kernel is selected for the SVMs and its optimum kernel width, i.e., γ in (1) as well as the trade-off parameter C are found by a grid search on the training set at each specific texton size and k value. To avoid too much computational cost for this grid search, 5-fold cross-validation at patient level (instead of leave-one-subject-out) is performed on the training set. This means that the training set is divided into five folds at patient level. One fold is used as the validation set and the remaining as the training set. Since the codebook has to be only constructed on the training set, we need to construct the codebook each time on the four folds used at this cross validation. We have, thus, repeated the experiments 10 times and averaged the results as there is a variation in the patches extracted each time for the construction of codebook.

4 Results

In this section, we first present the results for texton-based texture classification system using one single SVM as classifier with the parameters chosen as explained in previous section. Then the results of aggregation over different values of k with fixed texton-size and also aggregation over different texton sizes with fixed k are presented. Eventually, the comparison between single SVM and multiple classifier system is presented followed by the comparison with other techniques reported in the literature.

The results for using one single SVM are shown in Table 1 for various texton sizes and k values in k-means. The last row and last column on this table show the results of

using multiple classifier systems with product combiner that aggregates the decisions of single SVMs at various texton sizes or k values, respectively. These results are also shown graphically in Fig. 3 to make the comparison between single SVM and multiple classifier systems easier. As can be seen from top graph in Fig. 3, aggregation over various k values almost always yields better results than single k at the corresponding texton size. However, the bottom graph in Fig. 3 reveals that combining over various texton-sizes at the same k value does not produce better results than the best single SVM.

Comparison with Other Techniques. The comparison is made between the proposed texton-based classification system using multiple classifier systems with aggregation over k values and the results published in [21]. Since the same data as in [21] is used in our experiments, the results are directly comparable. In [21], the results are provided for several approaches among which we consider a filter bank approach using moments of histograms and an approach based on the LBP operators as follows:

1) GFB1 (Gaussian filter bank 1): using the moments of histogram computed on the outputs of convolved Gaussian filter banks with four rotation invariant filters obtained from linear combination of Gaussian derivatives at five scales.
2) LBP2: joint 2D LBP and intensity histograms.

The reader may refer to [21] for more information on these two approaches and also for further comparison with other techniques described therein. Moments of histograms computed on the outputs of Gaussian derivatives are one of the most common approaches in the literature for the classification of CT images of lung [2]. On the other hand, LBP2 reaches the best results among others in [21]. The results based on the above techniques are provided in Table 2 along with the best result obtained from the proposed approach based on texton signatures and multiple classifier system with aggregation over different k values using product combiner.

The confusion matrices for LBP2 and our best results are provided in Table 2. The proposed approach attains performance similar to LBP2 and McNemar's test also does not indicate significant difference ($p = 0.75$). The specificity of texton-based using MCS and LBP2 approaches are 96.61% and 93.33%, while their sensitivity are 95.37% and 97.25%, respectively (when comparing NT versus CLE and PSE).

Table 1. The results of texton-based classification system on CT images of lung used in this paper for k values of 10 to 50 and texton sizes (TS) of 3×3 to 9×9 pixels using a single SVM

Texton Size	$k = 10$	$k = 20$	$k = 30$	$k = 40$	$k = 50$	Aggregation over k
3 × 3	93.5 ± 1.1	93.2 ± 1.2	94.7 ± 1.4	93.0 ± 1.0	91.7 ± 1.1	94.5 ± 0.6
4 × 4	92.9 ± 0.7	93.6 ± 1.2	93.5 ± 1.3	94.1 ± 1.3	94.2 ± 0.9	**95.0 ± 0.6**
5 × 5	92.4 ± 1.0	91.7 ± 0.9	92.6 ± 0.9	92.7 ± 1.2	93.8 ± 1.3	94.2 ± 0.4
6 × 6	91.7 ± 1.3	90.8 ± 0.9	91.8 ± 1.5	90.3 ± 1.9	90.5 ± 1.4	92.1 ± 0.7
7 × 7	90.1 ± 1.4	91.1 ± 1.3	90.8 ± 1.0	89.8 ± 1.6	89.2 ± 1.7	91.1 ± 0.7
8 × 8	88.8 ± 1.7	89.5 ± 1.9	91.1 ± 0.9	91.0 ± 1.3	89.6 ± 1.8	91.7 ± 0.9
9 × 9	87.6 ± 1.4	88.8 ± 1.5	91.0 ± 1.1	89.8 ± 1.7	90.5 ± 1.0	90.8 ± 0.9
Aggregation over TS	92.1 ± 0.6	93.0 ± 0.7	93.5 ± 1.0	93.0 ± 0.6	92.8 ± 0.8	

Table 2. The comparison between the best results obtained from the proposed approach and the results of other techniques on the same data (*left*); the confusion matrix of LBP2 (*middle*) and texton-based approach with multiple classifier systems (MCS) and SVM as base classifier (*right*)

Technique	Accuracy
GFB1	61.3
LBP2	95.2
Texton-based using MCS	95.0

	Estimated Labels		
True Labels	NT	CLE	PSE
NT	55	0	4
CLE	1	49	0
PSE	2	1	56

	Estimated Labels		
True Labels	NT	CLE	PSE
NT	57	0	2
CLE	4	46	0
PSE	2	0	57

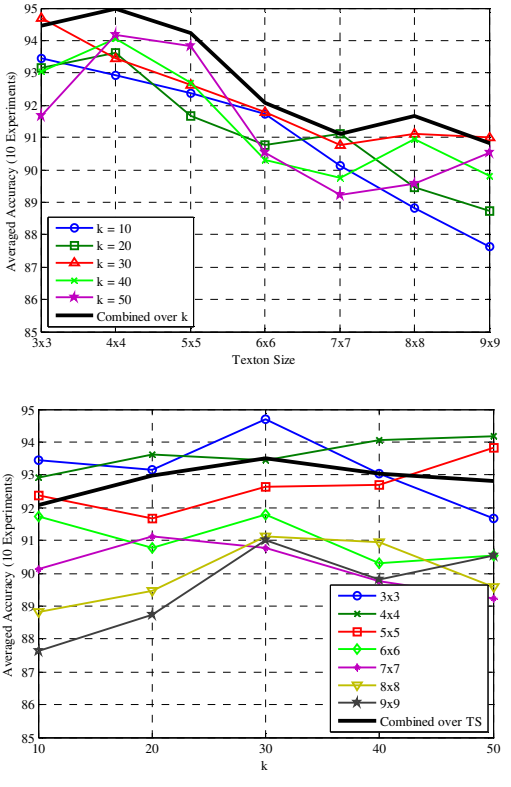


Fig. 3. The accuracy of the base classifiers and their combination on various texton sizes (TS) (*top*) and *k* values (*bottom*)

5 Discussions and Conclusion

In this paper, multiple classifier systems along with texton signatures are proposed for the classification of CT images of lung. Our results on the dataset of 168 ROIs of CT

images of lung shows that while texton-based approach using a single SVM has a satisfactory performance in this application; combining these single SVMs in a parallel structure over different k values slightly improves the classification results.

From Table 1 and top graph in Fig. 3, it seems that increasing texton size degrades the performance of single SVMs. This could be because larger texton sizes lead to higher dimensional space for k-means, requiring more data for reliable clustering. At the same time, there will be fewer sub-patches available with larger texton size for learning the model as described in Section 2.1. The aggregation results on various texton sizes show no improvement. A close look at the outputs of single SVMs at the same k value but different texton sizes reveals that most of the SVMs make the mistakes on the same ROIs and this means that there is a lack of diversity among the base classifiers. This explains why combining them do not improve the performance [11].

However, as can be seen from the bottom graph in Fig. 3 that displays the performance of single SVMs at a specific texton size at various k values, no certain value of k always yield the best results. At some texton sizes, larger k produces the best results (for example at texton size 4×4) while at other texton sizes medium or small k concludes best results. Since we do not know *a priori* which k produces the best results as the optimal k may vary from patient to patient depending on intrinsic scale and complexity of texture patterns, we aggregate over texton sizes at a specific k, which almost always produces better results than the best single SVM. By looking at the outputs of the base classifiers, it becomes clear that the diversity among them is higher than the previous case and this explains the improvement of the results by their combining.

Overall, we conclude that aggregating at smaller texton-sizes, for example 4×4 over different values of k reasonably produces good results which are similar to or better than the results obtained from other approaches in the literature. Among these approaches is LBP2 [21], which mainly relies on LBP operators.

Using parallel multiple classifiers systems on texton signatures proposed here can also be extended to LBP approach. LBP operators also involve two parameters, i.e., the size of operator (scale), and the number of bins in the estimation of histogram. The decisions based on single operators can be aggregated over any of these two parameters to investigate possible improvements.

In comparing LBP and texton-based approaches provided in this paper, one should notice that LBP operators are, by design, invariant to monotonic intensity transformations. While this is desirable in some applications, in the classification of Lung CT images, the mean of intensity is important and this justifies poor performance of an approach based on merely LBP operators as it discards the mean of intensity in the ROIs [21]. Due to this drawback of LBPs, in [20, 21], the joint intensity and LBP histograms are considered (LBP2). This improves the performance of the LBPs in this application at the cost of adding to the complexity of the approach. Texton-based approach does not suffer from this problem as it is not invariant to intensity transformations. On the other hand, LBP operators can be considered as fixed textons which are chosen irrespective of the data. Texton-based approach, however, extracts the textons using the data. This adds to the complexity of texton-based approach as the unsupervised step in learning the dictionary of textons is an extra step in this approach comparing to the LBPs. The performance of texton-based approach and LBP2 is similar for the data used in this paper. Nevertheless, our conjecture is that the superiority

of one approach to another is application dependent. If LBP operators can define a good representation for some data they conclude high performance while if texton-based approach can extract the textons accurately based on the training data, then it can yield high accuracy.

In future work, other classifiers that generate more diverse classification outputs than SVMs, such as decision trees, will be investigated. As mentioned above, in our experiments, SVMs as base classifiers lack diversity among themselves and, hence, combining their decisions does not yield significant improvement over best single SVM. We expect that decision trees, which are considered as ensemble of weak classifiers, generate more diverse outputs and their combination in a decision forest may conclude more improvement [22].

Also, in this paper, raw pixel representation is only used in texton-based approach. In some computer vision applications, it has been shown that building textons on the output of filter banks produces better accuracy [23, 24]. Although using raw pixel representation is computationally more attractive than using filter bank representation (as the intermediate step of convolving patches with the filter banks is not required), using filter banks in texton-based approach will be investigated in the future work for possible improvement of the results.

Acknowledgments. The funding from the Natural Sciences and Engineering Research Council (NSERC) of Canada under Canada Graduate Scholarship (CGS D3-378361-2009) and Michael Smith Foreign Study Supplements (MSFSS) is gratefully acknowledged.

References

1. Uppaluri, R., Mitsa, T., Sonka, M., Hoffman, E.A., McLennan, G.: Quantification of Pulmonary Emphysema from Lung Computed Tomography Images. Amer. J. Respir. Crit. Care Med. 156(1), 248–254 (1997)
2. Sluimer, I.C., Prokop, M., Hartmann, I., van Ginneken, B.: Automated Classification of Hyperlucency, Fibrosis, Ground Glass, Solid, and Focal Lesions in High-Resolution CT of the Lung. Medical Physics 33(7), 2610–2620 (2006)
3. Chabat, F., Yang, G.Z., Hansell, D.M.: Obstructive Lung Diseases: Texture Classification for Differentiation at CT. Radiology 228(3), 871–877 (2003)
4. Xu, Y., Sonka, M., McLennan, G., Guo, J., Hoffman, E.A.: MDCT-based 3-D Texture Classification of Emphysema and Early Smoking Related Lung Pathologies. IEEE Trans. Med. Imag. 25(4), 464–475 (2006)
5. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. IEEE Trans. Pattern Analysis and Machine Intelligence 24(7), 971–987 (2002)
6. Varma, M., Zisserman, A.: A Statistical Approach to Material Classification Using Image Patch Exemplars. IEEE Trans. Pattern Analysis and Machine Intelligence 31(11), 2032–2047 (2009)
7. Caputo, B., Hayman, E., Fritz, M., Eklundh, J.O.: Classifying Materials in the Real World. Image and Vision Computing 28(1), 150–163 (2010)

8. Varma, M., Zisserman, A.: A Statistical Approach to Texture Classification from Single Images. International Journal of Computer Vision: Special Issue on Texture Analysis and Synthesis 62(1-2), 61–81 (2005)
9. Gangeh, M.J., Sørensen, L., Shaker, S.B., Kamel, M.S., de Bruijne, M., Loog, M.: A Texton-Based Approach for the Classification of Lung Parenchyma in CT Images. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6363, pp. 596–603. Springer, Heidelberg (2010)
10. Garcia, M.A., Puig, D.: Supervised Texture Classification by Integration of Multiple Texture Methods and Evaluation Windows. Image and Vision Computing 25(7), 1091–1106 (2007)
11. Kuncheva, L.I.: Combining Pattern Classifiers Methods and Algorithms. John Wiley & Sons, New Jersey (2004)
12. Lanckriet, G.R.G., Cristianini, N., Bartlett, P., Ghaoui, L.E., Jordan, M.I.: Learning the Kernel Matrix with Semidefinite Programming. Journal of Machine Learning Research 5(1), 27–72 (2005)
13. Bach, F.R., Lanckriet, G.R.G., Jordan, M.I.: Multiple Kernel Learning, Conic Duality, and the SMO Algorithm. In: Proceedings of 21st International Conference of Machine Learning, ICML (2004)
14. Julesz, B.: Textons, the Elements of Texture Perception, and Their Interactions. Nature 290(5802), 91–97 (1981)
15. Leung, T., Malik, J.: Representing and Recognizing the Visual Appearance of Materials Using Three-Dimensional Textons. International Journal of Computer Vision 43(1), 29–44 (2001)
16. Cula, O.G., Dana, K.J.: 3D Texture Recognition Using Bidirectional Feature Histograms. International Journal of Computer Vision 59(1), 33–60 (2004)
17. Fan, R.E., Chen, P.H., Lin, C.J.: Working Set Selection Using the Second Order Information for Training SVM. Journal of Mach. Learning Research 6, 1889–1918 (2005)
18. Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On Combining Classifiers. IEEE Trans. Pattern Analysis and Machine Intelligence 20(3), 226–239 (1998)
19. Webb, W.R., Müller, N., Naidich, D.: High-Resolution CT of the Lung, 3rd edn. Lippincott Williams & Wilkins (2001)
20. Sørensen, L., Shaker, S.B., de Bruijne, M.: Texture Classification in Lung CT Using Local Binary Patterns. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 934–941. Springer, Heidelberg (2008)
21. Sørensen, L., Shaker, S.B., de Bruijne, M.: Quantitative Analysis of Pulmonary Emphysema Using Local Binary Patterns. IEEE Trans. Med. Imag. 29(2), 559–569 (2010)
22. Ho, T.K.: The Random Subspace Method for Constructing Decision Forests. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(8), 832–844 (1998)
23. Tuzel, O., Yang, L., Meer, P., Foran, D.J.: Classification of Hematologic Malignancies Using Texton Signatures. Pattern Analysis and Applications 10(4), 277–290 (2007)
24. Zhong, C., Sun, Z., Tan, T.: Robust 3D Face Recognition Using Learned Visual Codebook. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2007), pp. 1–6 (2007)

Imaging as a Surrogate for the Early Prediction and Assessment of Treatment Response through the Analysis of 4-D Texture Ensembles (ISEPARATE)

Peter Maday[1], Parmeshwar Khurd[1], Lance Ladic[1], Mitchell Schnall[2], Mark Rosen[2], Christos Davatzikos[2], and Ali Kamen[1]

[1] Siemens Corporate Research,
Princeton, NJ, USA
{Peter.Maday.ext,Parmeshwar.Khurd,
Ali.Kamen,Lance.Ladic}@siemens.com
[2] University of Pennsylvania,
Philadelphia, PA, USA
{Mitchell.Schnall,rosenmar,Christos.Davatzikos}@uphs.upenn.edu

Abstract. In order to facilitate the use of imaging as a surrogate endpoint for the early prediction and assessment of treatment response, we present a quantitative image analysis system to process the anatomical and functional images acquired over the course of treatment. The key features of our system are deformable registration, texture analysis via texton histograms, feature selection using the minimal-redundancy-maximal-relevance method, and classification using support vector machines. The objective of the proposed image analysis and machine learning methods in our system is to permit the identification of multi-parametric imaging phenotypic properties that have superior diagnostic and prognostic value as compared to currently used morphometric measurements. We evaluate our system for predicting treatment response of breast cancer patients undergoing neoadjuvant chemotherapy using a series of MRI acquisitions.

Keywords: therapy response, image registration, texture classification.

1 Introduction

Imaging is often used to predict and assess the response to a particular form of therapy or treatment [1]. However, the lack of adequate quantitative tools often forces this assessment to be performed in a qualitative manner. Existing tools based upon analysis of simple measures such as lesion volume or number cannot reliably predict whether the treatment is effective at an early stage because such measures ignore the location(s) of the lesion(s) and the detailed spatial characteristics of each lesion. Moreover, such tools cannot be used to analyze diseases that manifest themselves as spatially diffuse abnormalities as opposed to spatially compact lesions. Therefore, we propose a quantitative tool based upon deformable registration and texture analysis of the images acquired over time, followed by feature selection and classification.

B. Menze et al. (Eds.): MICCAI 2010 Workshop MCV, LNCS 6533, pp. 164–173, 2011.

The proposed framework is described in Section 2 and has been evaluated (please see Section 3) on a dataset containing acquisitions of breast cancer patients undergoing neoadjuvant chemotherapy. For the analysis, DCE (Dynamic Contrast Enhanced)-MRI images have been used [2]-[4].

2 Methods

Systematic evaluation of change that occurs with treatment and disease progression necessitates the ability to precisely measure change in the tissue parameters. To provide a basis for tissue comparison, the differences introduced by the variations in patient position between the pre-treatment baseline scan and the follow-up scans need to be eliminated. To reduce this spatial variability, the follow-up images are registered to the baseline acquisition of each individual patient in all cases. A region of interest (ROI) containing the abnormality is specified in the baseline reference frame. Holistic features representing the whole ROI are created by aggregating voxel-based features characterizing anatomical and functional properties. Texture features from anatomical imaging modalities help in characterizing the variability of the tissues. Features based on functional images help in understanding the physiological properties of the abnormalities that cannot be inferred from the anatomical images only.

For predicting the treatment outcome, a classifier is trained with the features extracted from the ROI. The features are first subjected to a feature selection technique in order to boost the classification performance. The acquisition times and the expert labeling of the patients indicating the clinical outcome are used for training and evaluation. Our entire system is displayed in Fig. 1. We now describe each system component in detail.

2.1 Registration

In order to compare changes in the corresponding regions on the images, a robust registration method is required that reduces the spatial variability among the selected regions. We first perform affine registration of the acquired images followed by deformable registration. This is done pair-wise between a follow-up image and the corresponding baseline image. In the deformable stage, we use a multi-resolution registration scheme, where the transformation degrees of freedom are increased as the registration process progresses from the initial to the final stages. To avoid the necessity of intensity normalization and inhomogeneity correction for the datasets, we have used the normalized mutual information [3, 5] metric as the similarity measure for the registration. During the iterations for a given resolution, we update the displacement field via gradient descent and perform regularization based on Gaussian filtering.

To accomplish local incompressibility, we select the regularization parameters in order to maintain a high level of spatial regularization. This brings the gross structures present in the images into alignment, but preserves the local texture in the ROI within the registered images. Alternative methods have been proposed utilizing regularization terms that penalize the deviation of the Jacobian determinant of the deformation fields from unity [3]. We have evaluated the volume-preserving properties of our method by considering the Jacobian determinants of the resulting deformation fields

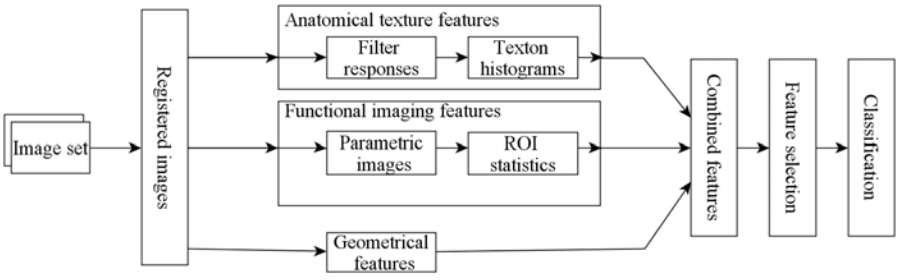


Fig. 1. The stages of evaluation within the proposed system. Special care needs to be taken for the creation of combined features (see Section 2.3).

and did not find evidence of local tissue compression with Jacobian below unity. Moreover, we note that the texture analysis carried out at later stages (see Section 2.2) is chosen to be robust to minor artifacts in the registration.

2.2 Features Extracted from Anatomical and Functional Images

Measurement of basic tumor parameters such as volume are commonly used to quantify tumor characteristics and treatment response. However, the large variability in the morphological phenotype of tumors indicates the need for features describing low-level texture characteristics. We have used Gabor filters, which are a family of multi- scale texture features commonly used in computer vision applications. The two dimensional Gabor function $g(x, y)$ is given in the following form:

$$g(x, y) = \left(\frac{1}{2\pi\sigma_x\sigma_y}\right)\exp\left[-\frac{1}{2}\left(\frac{x}{\sigma_x^2}+\frac{y}{\sigma_y^2}\right)+2\pi jWx\right]$$

where σ_x, σ_y are filter scaling parameters and W is a frequency shift parameter. A class of functions is formed by dilations and rotations of the generating function above:

$$g_{mn}(x, y) = a^{-m}g(x', y') \quad a>1, m, n\in Z$$
$$x' = a^{-m}(x\cos\theta + y\sin\theta), \quad y' = a^{-m}(-x\cos\theta + y\sin\theta)$$

where $\theta = n\pi / K$ and K is the total number of orientations and a is a scale normalization parameter.

The texture features are obtained by filtering the images with a set of Gabor functions using different parameter settings. Following the approach suggested in [6], the selection of parameters is conducted in the frequency domain in order to compute filter responses that uniquely capture certain textural properties of the filtered images. To achieve rotational invariance, the maximum value is selected for each scale over all orientations.

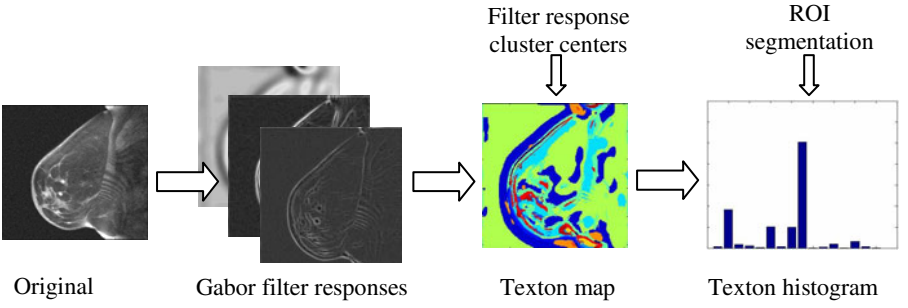


Fig. 2. From left to right; the intermediate results used for the construction of texton histograms are shown on breast MRI images

To improve descriptive performance and provide robustness against minor errors in the registration, histograms of vector quantized filter responses computed over the ROI are used in the following stage as shown in Fig. 2. From a set of training samples, clusters are formed and the cluster centers or *textons* [7] are used as prototypes of filter responses. Given an image, the computed filter outputs are assigned to the nearest texton based on the Euclidean distance [7]. A histogram is then formed representing the distribution of texton assignments in the specified region of interest. Texton histograms have been successfully applied for the classification of texture samples in the literature [7,8].

Although changes related to the treatment response in the imaged regions may not be immediately visible on the anatomical images, functional imaging (DCE-MRI, PET, etc.) often proves useful in the prediction of response by providing insight into the physiological condition of the tissues. In DCE-MRI, quantities related to the diffusion of contrast materials can be estimated by fitting a pharmacokinetic model [9] on the time course of the contrast concentration for each voxel resulting in parametric images. As little is known about the relevance of local differences in the spatio- temporal characteristics of contrast uptake, we have restricted our attention to the use of statistical quantities (mean, std. deviation, kurtosis) of parametric images computed over the ROI. Such an approach has previously been proposed in [10].

2.3 Comparing and Combining Features from Differing Acquisition Times

In order to provide a prediction about the efficacy of the treatment, we need to compare patients by combining the feature sets across acquisitions in a uniform and consistent manner. Note that the baseline image obtained prior to commencement of the treatment already contains information about whether a treatment might work.

Moreover, the time distribution of the follow-up scans may not be identical for each patient, and a significant overlap might be present in the distributions of the acquisition times (see Fig. 3.). We expect a reasonable level of coherence to be present in the acquisition times based upon the specific acquisition protocol followed. For this reason, we create virtual time groups of the acquisitions by finding clusters in the scanning dates. By assuming that the changes experienced as an effect of the treatment have a consistent temporal behavior, we assign the acquisitions to the group with the closest center.

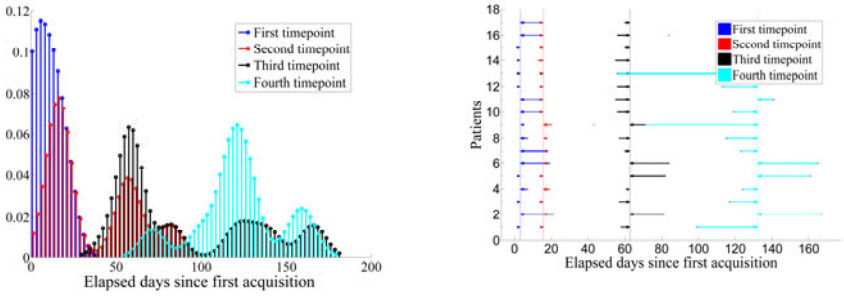


Fig. 3. The left image presents the distribution of the number of days elapsed since the first baseline image acquisition. The distributions are shown for each time point in the acquisition series with different colors. The right image shows the results of the time point assignment to the groups determined by the four cluster centers (vertical lines).

A K-means algorithm was used to perform clustering of acquisition dates [11]. The number of clusters has been initialized to the most common number of image acquisitions for the patients. The concept is illustrated in Fig. 3. The blue, red, black and cyan lines on the right image in Fig. 3 indicate the virtual cluster assignments for the choice of 4 clusters. The actual data-points themselves have been shown with small black crosses. Note that the image at day 55 for patient 13 is assigned to both clusters 3 and 4 since this patient had only 3 follow-up scans and the final follow-up was close to the last cluster center.

Within each of the acquisition date time clusters, the state of the disease is comparable due to the availability of the same feature values, and the temporal proximity resulting from the assignment. For every patient, we first form acquisition-specific feature vectors for the baseline scan and each virtual time cluster by consistently concatenating the texton features, kinetic features and geometric features extracted from the corresponding anatomical and functional images. We then order these feature vectors by their cluster-center times and concatenate them to obtain a larger patient-specific feature vector.

2.4 Feature Selection and Classification

Classifier performance is negatively impacted in situations where the number of features is large compared to the size of the available training set. To address this, a Minimal-Redundancy-Maximal-Relevance (MRMR) feature selection method was used to remove unnecessary features [12]. The method aims to find a subset S of features, with the maximal mutual information to the class labels, while trying to avoid redundancy between the features themselves. If *I(M,N)* denotes the mutual information [11] between any two random variables *M* and *N*, then MRMR is defined as follows:

$$\max_S \Phi = D - R \text{, where} \quad D = \frac{1}{|S|}\sum_{x_i \in S} I(x_i, c) \quad R = \frac{1}{|S|^2}\sum_{x_i, x_j \in S} I(x_i, x_j)$$

The quantity Φ is maximized over the set S containing the indices of the selected features, and the variable x_i denotes the i-th feature and c denotes the class labels. The optimization using the MRMR method can be efficiently performed in an incremental manner by extending the set of selected features by one candidate at a time, while guaranteeing first-order optimality with respect to the maximum dependency criterion [12].

For the classification of feature vectors, the Support Vector Machine (SVM) [13] algorithm was used. The SVM hyper-parameters were evaluated by a grid search in the parameter space and its classification accuracy has been determined by cross validation.

3 Results

The proposed system has been evaluated on a set of 17 patients with breast cancer. Patients have been undergoing neoadjuvant chemotherapy, and were imaged on a regular basis throughout the course of the treatment. The cases were assigned to two groups, responders and non-responders, after the end of the treatment by medical experts on the basis of non-imaging biochemical serum tests. The dataset contains 8 patients labeled as responders, and 9 patients as non-responders. For each acquisition time point, a series of 3-4 DCE-MRI images was obtained. The dimensions of the images were 256x256x64 with a voxel size of 0.7mm x 0.7mm x 2mm. The first image of the set has been taken before the injection was given, and the rest were taken after, with a regular time interval of 5 minutes in between each image acquisition. In most of the cases, the image acquisition set consists of three images.

For each patient, the non-contrast-enhanced images for subsequent timepoints were non-rigidly registered to the baseline acquisition using 4 resolutions in the method described in Section 2.1 (please refer to Fig. 4 for an example). To avoid unnecessary complications arising from registering DCE-MRI images, we used the deformation fields obtained by the registration of the non-contrast enhanced images for their alignment, as the body position differences between the DCE frames are negligible.

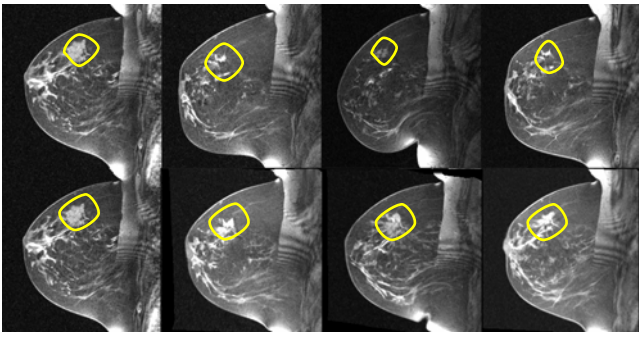
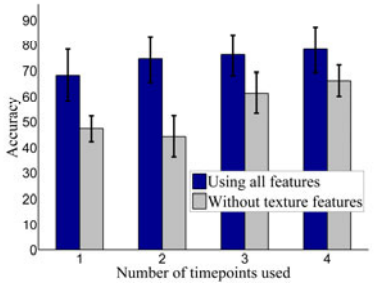


Fig. 4. The top row shows non contrast enhanced images for a patient acquired over the course of the treatment. The bottom row contains the registration results for the same images. The region containing the tumor is shown highlighted. The bottom row clearly demonstrates the above patient to be a non-responder.



Timepoints used for classification	Classification accuracy without texture features	Classification accuracy using all features
Baseline only	47.41. ± 5 %	68.35 ± 10 %
+ 1 followup	44.65 ± 8 %	74.59 ± 9 %
+ 2 followups	61.42 ± 8.%	76.24 ± 8 %
+ 3 followups	66.22 ± 6.%	78.35 ± 9 %

Fig. 5. The left image illustrates the classification accuracy of our system with a varying number of timepoints. The table on the right hand side contains the numerical results obtained with our system (third column) and a more conventional approach (second column) without any texture features or feature selection.

To manage inconsistencies in the intensity values between acquisitions, linear normalization has been applied to the image sets as a preprocessing step during feature extraction. Due to the diffuse nature of the tumors, the region containing a compact portion with high contrast intensity was manually selected as the ROI. The tumor volume was measured by segmentation on the difference image of the peak contrast and non-contrast acquisitions. A Gabor filter bank consisting of a set of filters with 6 orientations and 4 scales was used for the extraction of texture features. The real and imaginary parts of the filter responses were used independently. To better characterize the volumetric properties of the tissues, the filter responses for every location were evaluated along two perpendicular planes and the responses were treated as independent features. We used 20 texton histogram bins. The limited number of acquired DCE-MRI images (3-4) did not permit the fitting of a generic pharmacokinetic model. Therefore, we evaluated the statistical measures on parametric images constructed using the slopes of linear approximations of the rise and decay stages of the contrast uptake as the parameters representing the micro-vascular structure of the tumors. These parameters were also used in [4] for the classification of benign versus malignant lesions and related parameters such as the micro-vascular vessel wall permeability (K^{trans}) and extracellular volume fraction (ν_e) have been used in [14]. In the end, each baseline/follow-up virtual timepoint had 27 features: 20 texton histogram counts, 6 kinetic features: mean, variance, kurtosis for rise/decay images, and 1 geometric feature: tumor volume.

To measure the usefulness of the system for early prediction of response, we evaluated the classification accuracies using subsets of the dataset. We first predict treatment response using only the baseline scan. In addition, we used the acquisition date time clustering from Section 2.3 with K=3 to obtain classification accuracies using 1, 2 and 3 virtual follow-up time-points. The obtained classification accuracies in Fig. 5 show the advantages of including acquisitions from later time-points and the use of texture features along with feature selection. For each of these 4 cases, the 20 highest ranking features were selected. The mean classification accuracy and its

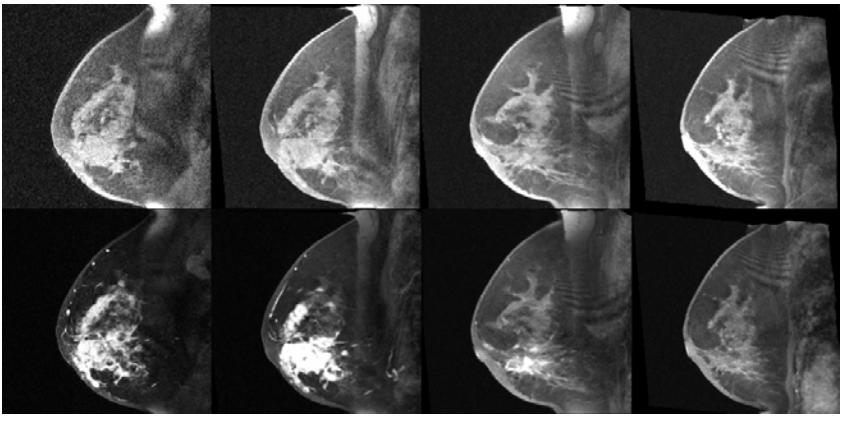


Fig. 6. The top row contains registered images of respective slices for the same patient throughout the acquisitions. The bottom row shows the same slices with contrast agent injection. Even though only minor changes are observable in the final distribution of the lesion on the non-contrast enhanced MRI image, the patient is a confirmed responder. The contrast-enhanced images, however, show a change in the angiogenesis.

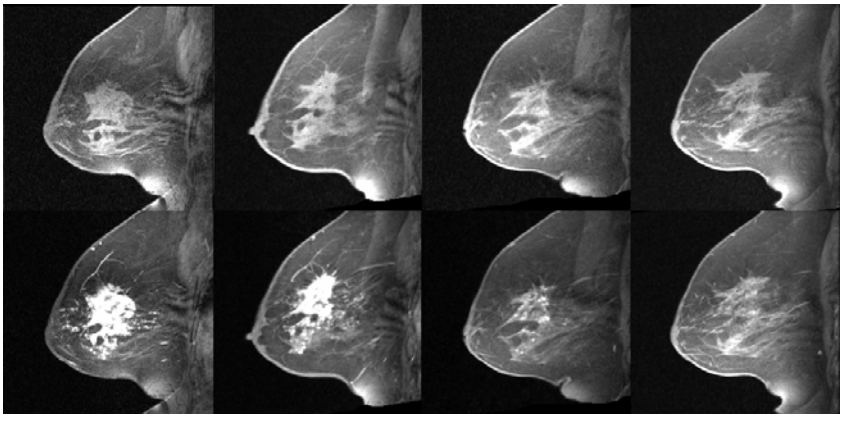


Fig. 7. The top row contains registered images of respective slices for the same patient throughout the acquisitions. The bottom row shows the same slices with contrast agent injection. Although being a non-complete responder, easily perceivable decrease of tumor size is present in the contrast-enhanced images.

standard error were obtained using 10-fold cross validation and repeating the experiment 50 times to eliminate the uncertainties caused by the assignment of the samples to the folds. We used an SVM with a radial basis function (RBF) kernel. The MRMR feature selection was performed on each training fold during 10-fold cross-validation. The selection of hyper-parameters (i.e., the RBF and the SVM slackness parameters) was also conducted via internal cross-validation on the training fold with a grid search using an exponential scale. In order to study sample-size effects, the proposed feature

selection and classification approach was additionally evaluated by selecting 17 samples from standard datasets such as the arrhythmia dataset belonging to the UCI repository. The observed 10-fold classification accuracy on the 17-sample dataset was about 5-10% lower than the 10-fold classification accuracy on the entire dataset.

In order to understand the reason behind our low classification accuracy of 78% in Fig. 5 despite the use of deformable registration and texture analysis, we conducted a thorough visual analysis of our images. Through figures 6-7, we demonstrate the inherent complexity in designing an image classification system for our dataset. In Fig. 6, we show that a responding patient who shows minor changes in the non-contrast-enhanced images over time, but major changes in the contrast-enhanced images over time. However, in Fig. 7, we show that a non-responder can exhibit a similar behavior. The images indicate that complete recovery may or may not occur despite some treatment response. Since this is an emerging application area, the quantitative accuracy of radiologists in assessing or predicting treatment response via DCE-MRI is not exactly known.

4 Discussion and Conclusion

We have introduced a quantitative image analysis system for predicting and assessing treatment response. Furthermore, we have demonstrated the efficacy of our system in predicting the response of breast cancer patients undergoing neoadjuvant chemotherapy by evaluating our system on a dataset comprising of 17 patients with breast cancer. We note that since baseline images obtained prior to commencing treatment contain predictive information about treatment effectiveness, our system could pave the way for personalized medicine. This paper presents preliminary results and is about the method and its promise in this emerging application area. In future work, we plan to evaluate our system on a larger cohort of patients. Other potential applications of our method include the evaluation of the "watchful waiting" treatment option for prostate cancer management via DCE-MRI.

References

1. Pathak, S.D., Ng, L., Wyman, B., Fogarasi, S., Racki, S., Oelund, J.C., Sparks, B., Chalana, V.: Quantitative image analysis: software systems in drug development trials. Drug Discovery Today 8(10), 451–458 (2003)
2. Lorenzon, M., Zuiani, C., Londero, V., Linda, A., Furlan, A., Bazzocchi, M.: Assessment of breast cancer response to neoadjuvant chemotherapy: Is volumetric MRI a reliable tool?. European Journal of Radiology 71(1), 82–88 (2009)
3. Li, X., Dawant, B.M., Brian Welch, E., Bapsi Chakravarthy, A., Freehardt, D., Mayer, I., Kelley, M., Meszoely, I., Gore, J.C., Yankeelov, T.E.: A nonrigid registration algorithm for longitudinal breast MR images and the analysis of breast tumor response. Magnetic Resonance Imaging 27(9), 1258–1270 (2009)
4. Zheng, Y., Baloch, S., Englander, S., Schnall, M.D., Shen, D.: Segmentation and Classification of Breast Tumor Using Dynamic Contrast-Enhanced MR Images. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 393–401. Springer, Heidelberg (2007)

5. Studholme, C., et al.: An overlap invariant entropy measure of 3D medical image alignment. Pattern recognition 32(1), 71–86 (1999)
6. Manjunath, B.S., Ma, W.Y.: Texture features for browsing and retrieval of image data. IEEE Transactions on Pattern Analysis and Machine Intelligence 18(8), 837–842 (1996)
7. Varma, M., Zisserman, A.: Classifying images of materials: Achieving viewpoint and illumination independence. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2352, pp. 255–271. Springer, Heidelberg (2002)
8. Zhang, J., Marszalek, M., Lazebnik, S., Schmid, C.: Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study. International Journal of Computer Vision 73(2), 213–238 (2007)
9. Tofts, P.S., Brix, G., Buckley, D.L., Evelhoch, J.L., Henderson, E., Knopp, M.V., et al.: Estimating kinetic parameters from dynamic contrast-enhanced T1-weighted MRI of a diffusible tracer: standardized quantities and symbols. J. Magn. Reson. Imaging 10, 223–232 (1999)
10. Chang, Y.-C., Huang, C.-S., Liu, Y.-J., Chen, J.-H., Lu, Y.-S., Tseng, W.-Y.I.: Angiogenic response of locally advanced breast cancer to neoadjuvant chemotherapy evaluated with parametric histogram from dynamic contrast-enhanced MRI. Physics in Medicine and Biology 49(16), 3593–3602 (2004)
11. Bishop, C.M.: Pattern recognition and machine learning. Springer, New York (2006)
12. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(8), 1226–1238 (2005)
13. Chang, C.-C., Lin, C.-J.: LIBSVM: a library for support vector machines (2001), Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`
14. Yankeelov, T.E., Luci, J.J., Lepage, M., Li, R., Debusk, L., Charles Lin, P., Price, R.R., Gore, J.C.: Quantitative pharmacokinetic analysis of DCE-MRI data without an arterial input function: a reference region model. Magnetic Resonance Imaging 23(4), 519–529 (2005)

A Texture Manifold for Curve-Based Morphometry of the Cerebral Cortex

Maxime Boucher[1,⋆], Alan Evans[2], and Kaleem Siddiqi[1]

[1] Center for Intelligent Machines, McGill University
{boucher,siddqi}@cim.mcgill.ca
[2] McConnell Brain Imaging Center, McGill University
alan@bic.mni.mcgill.ca

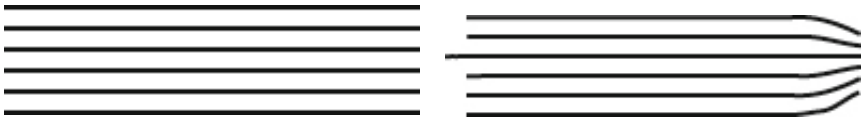
Abstract. The cortical surface of the human brain is composed of folds that are juxtaposed alongside one another. Several methods have been proposed to study the shape of these folds, e.g., by first segmenting them on the cortical surface or by analysis via a continuous deformation of a common template. A major disadvantage of these methods is that, while they can localize shape differences, they cannot easily identify the directions in which they occur. The type of deformation that causes a fold to change in length is quite different from that which causes it to change in width. Furthermore, these two deformations may have a completely different biological interpretation. In this article we propose a method to analyze such deformations using directional filters locally adapted to the geometry of the folding pattern. Motivated by the texture flow literature in computer vision we recover flow fields that maintain a fixed angle with the orientation of folds, over a significant spatial extent. We then trace the flow fields to determine which correspond to the shape changes that are the most salient. Using the OASIS database, we demonstrate that in addition to known regions of atrophy, our method can find subtle but statistically significant shape deformations.

1 Introduction

The human cortical surface is composed of a set of folds that run alongside one another to form an undulating pattern of crests and troughs. The shape of this folding pattern evolves during the normal cycle of human brain development under the effect of aging or the presence of diseases or cognitive deficits. A popular method to capture shape differences is to view brain development as a continuous deformation of a common template [1,2,3]. The continuous deformation between the common template and individual cortical surfaces is typically constrained to be a diffeomorphism and is obtained using surface registration techniques. Deformation statistics are then used to find regions of significant tissue growth or atrophy. Such statistics can determine, for example, if the overall area of a particular region is larger in one group compared with another.

⋆ Corresponding author.

B. Menze et al. (Eds.): MICCAI 2010 Workshop MCV, LNCS 6533, pp. 174–183, 2011.



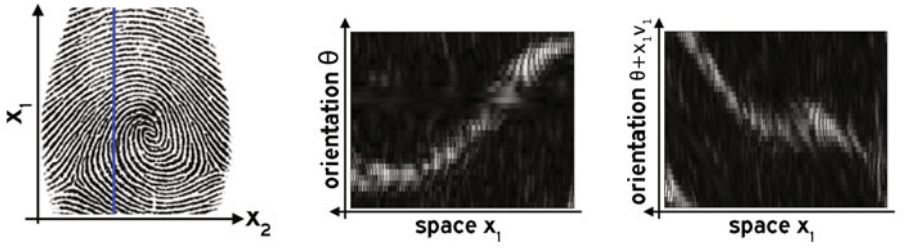
(a) (b)

Fig. 1. Fig. 1(b) shows two deformations of the folding pattern in Fig. 1(a). On Fig. 1(b), a deformation parallel to a fold increases its length, and perpendicular to folds decreases the spacing between folds.

In addition to diffeomorphism based statistics, which capture local changes in shape, regularization filters such as the isotropic diffusion kernel on surfaces [4,5], can yield a more global measure of shape differences. However, anatomical structures in the human brain are typically not isotropic and nor are the changes they induce on the cortical surface as they deform. As an example Fig. 1 illustrates two distinct deformations of a folding pattern, represented by a set of parallel curves. In the context of cortical shape analysis it is therefore beneficial to analyze differences using filters that are tuned to specific orientations. A neuroscientist can then examine whether an observed shape difference in a group follows a particular direction relative to fold orientation.

In this paper we abstract the shape of cortical folds as a collection of juxtaposed curves that locally have similar orientations. We then design shape filters which maintain a constant angle with neighboring folds. For example, a particular shape filter can be designed to be sensitive to a deformation parallel to fold orientation (an angle of 0 degrees), while another can be sensitive to a deformation that is perpendicular. This permits an analysis which is relative to cortical fold orientation. The challenge here is the estimation of the curvature to apply so that each filter maintains this constant angle.

To illustrate the computation of curvature consider the fingerprint example of Fig. 2(a), which is a pattern formed of juxtaposed intensity ridges similar to the folds of the cerebral cortex. The output of a directional ridge detector applied to the slice in blue is shown in Fig. 2(b), as an intensity function of angle (θ) versus position x_1. It is evident that this output aggregates along a a continuous curve (or submanifold), dubbled a texture flow in the work of [6]. When the exact location of the submanifold of maximum intensity in Fig. 2(b) is known the submanifold can be characterized as a function of orientation in space $\Theta(x_1, x_2)$. Here Θ gives the orientation of neighboring curves at a location (x_1, x_2) of Fig. 2(a). The curvature of neighboring curves is expressed as the gradient of Θ as $\nabla\Theta$. On the other hand, the slope of the aggregated intensity curve in Fig. 2(b) is also given by $\nabla\Theta$. Our aim in this paper is to determine the slope of the texture flow, and then generate directional filters using the curvature implied by this slope.



(a) Fingerprint image

(b) Ridge detection on Fig. 2(a) along the blue slice

(c) Fig. 2(b) resampled along a given slope v_1.

Fig. 2. A texture manifold depicts change in orientation of the neighboring curves in space

Inspired by the work of [6] we achieve a globally optimal assignment of curvatures to the cortical surface by using a dictionary of smooth flow fields on the surface. Each smooth flow field provides a hypothesis for the slope of the texture flow on the cerebral cortex. We measure the goodness of fit between each flow field and the actual texture manifold by using anisotropic filters. We then use relaxation labeling to achieve a globally optimum assignment of smooth flow fields that match the cortical folds on the surface. Directional filters are then generated by launching streamlines whose curvature is provided by this assignment.

Using the OASIS database [7] we test our framework to determine how the shape of cortical folds in patients with cognitive impairment is affected. We are able to identify folds that undergo significant deformation in the temporal lobe, and also a stretch in length of the Cingulate gyrus below the Praecuneus area and below the sup frontal cortex in the left hemisphere. These results are also partially revealed by a statistical analysis based on surface area (see Fig. 5). However, our results clearly indicate that the increase in surface area comes from a stretch in sulcal length (see Fig. 6(b)). Whether this stretch in length is due to structural connectivity loss in the white matter remains to be investigated, but our findings are corroborated by a visual inspection of shape differences between the two groups.

2 Estimating the Texture Manifold from Folds

We give a brief technical overview of the algorithm before detailing each step. We refer again to Figure 2 to illustrate the method. Suppose that the output of the ridge detection in Figure 2(a) can be expressed as an intensity function $I(x_1, x_2, \theta) \rightarrow \mathbb{R}^+$. We formulate different hypotheses H_i about the slope of the texture flow. Suppose, for example, that one such hypothesis H_i states that the texture flow has a slope of (v_1, v_2). This means that if we resample I as

$$I_i(x_1, x_2, \theta) = I(x_1, x_2, \theta + v_1 x_1 + v_2 x_2), \tag{1}$$

then, if this hypothesis is good, the intensity function should aggregates along an "horizontal" plane (as shown in Fig. 2(c)). A simple way to measure if a particular hypothesis is a good explanation for the observed texture flow is to measure how horizontal it is using an anisotropic filter. Suppose that $g_{a,b}$ is a Gaussian kernel where a expresses the width of the kernel along the spatial dimensions (x_1, x_2) and b expresses the width of the kernel along the orientation dimension θ as

$$g_{a,b}(x_1, x_2, \theta) = \left((2\pi)^{3/2} a^2 b\right)^{-1} exp\left(-\frac{x_1^2 + x_2^2}{2a^2} - \frac{\theta^2}{2b^2}\right). \tag{2}$$

A good measure to determine if H_i is a good hypothesis is

$$f_i(x_1, x_2, \theta) = \left(\frac{\partial}{\partial\theta} I_i * g_{a,b}\right)^2 \tag{3}$$

where $*$ is a convolution. The best hypothesis H_i is the one that locally maximizes Equation 3. However, an optimal hypothesis should be a good fit for several curves in a large neighborhood. We therefore use each hypothesis H_i as labels in the space formed by (x_1, x_2, θ). We achieve a globally optimum labeling $H^*(x_1, x_2, \theta)$ by selecting, at each point, the label which is both a good fit to the intensity function I and is the most similar to the neighboring labels.

The ridge detector for surfaces is presented in Section 2.1, the algorithm to generate flow hypotheses H_i on surfaces is explained in Section 2.2 and the relaxation labeling approach to find an optimal assignment is explained in Section 2.3. Once an optimal assignment is reached, it is possible to generate streamlines that follow the geometry of the folding pattern, which can then be used to detect shape changes. The algorithm to generate streamlines and use them to test statistics on surfaces is explained in Section 2.4.

2.1 A Ridge Detector on Surfaces

In this paper, we used the principal curvature of the surface to detect ridges and generate an intensity function I that describes the texture manifold. We present the intensity function I that gave the best result, however the algorithm applies for other choices of ridge detectors as well.

Let $\mathcal{S}$ be a smooth genus 0 surface and let $\kappa_1, \kappa_2, |\kappa_1| \geq |\kappa_2|$ be the principal curvatures of $\mathcal{S}$ and $\mathbf{u}_1, \mathbf{u}_2$ their associated vector directions. We associate a fold with a line of low curvature in one direction with high curvature in the perpendicular direction. Let $\mathbb{S}$ be the unit circle and let $\mathbf{v}_\theta \in \mathbb{S}$. We define the probability to observe a fold at a given location $\mathbf{u} \in \mathcal{S}$ for a given orientation θ using a symmetric Von Mises distribution as

$$I(\mathbf{u}, \theta) = n(|\kappa_1(\mathbf{u})| - |\kappa_2(\mathbf{u})|)^{-1} e^{(|\kappa_1(\mathbf{u})| - |\kappa_2(\mathbf{u})|)(\mathbf{u}_\theta^t \mathbf{u}_1)^2}, \tag{4}$$

where $n(|\kappa_1(\mathbf{u})| - |\kappa_2(\mathbf{u})|)$ is the normalization factor such that $\int_{\mathbb{S}} I(\mathbf{u}, \theta) d\theta = 1$. Equation 4 can be seen as the equivalent of a Gaussian distribution for angles, with a maximum at $\mathbf{u}_\theta = \pm\mathbf{u}_1$ and where $|\kappa_1| - |\kappa_2|$ determines the spread around the maximums.

2.2 A Dictionary of Smooth Vector Fields to Model the Texture Manifold

Unlike in images, it is not possible to hypothesize that the texture flow has a slope given by v_1, v_2 and then resample the image using these slope parameters as done in Equation 1. Instead, we use a base flow field $\mathbf{h}_i$. Let $\mathbf{h}_i(\mathbf{u})$ be the angle of the base flow field at a point $\mathbf{u} \in \mathcal{S}$. We can use $\mathbf{h}_i$ to resample I as

$$I_i(\mathbf{u}, \theta) = I(\mathbf{u}, \theta + \mathbf{h}_i(\mathbf{u})) \ . \tag{5}$$

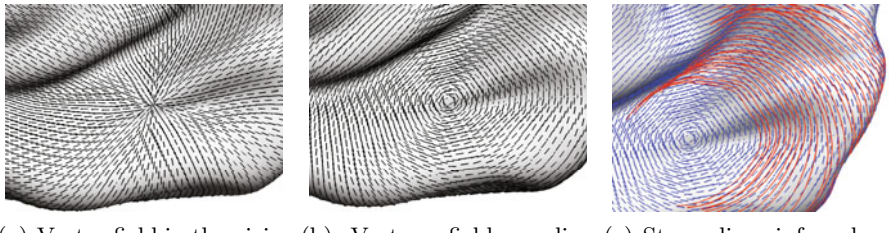
Then, we can apply a ridge detection to determine if $\mathbf{h}_i$ is a good hypothesis locally of the texture manifold.

The point of the algorithm is to generate several flow field $\mathbf{h}_i$, each with a different curvature. Each flow field $\mathbf{h}_i$ is generated by placing a source and a sink on $\mathcal{S}$ and then generating a smooth (singularity-free) completion between them. Specifically, let $\mathbf{s}_i = (\mathbf{s}_{i,1}, \mathbf{s}_{i,2})$ be a source and a sink, respectively. A vector field is then completed on $\mathcal{S}/\mathbf{s}_i$ by finding the minimum of the following functional:

$$\mathbf{h}_i = argmin_{\mathbf{h}^*} \int_{\mathcal{S}/\mathbf{s}_i} curl(\mathbf{h}^*)^2 + div(\mathbf{h}^*)^2 d\mathcal{S}. \tag{6}$$

We note that $\mathbf{h}_i$ is a unit vector field defined on $\mathcal{S}$. An example for a $\mathbf{h}_i$ is given in Figure 3(a). On this figure, the vector field fans in the vicinity of the singularity. To generate a full set of hypothesis $\mathbf{h}_i, i = 1, \ldots, N$, we distribute sources and sinks uniformly on the cortical surface such that every location is offered a possibility to "fan" . An interesting property of our formulation in Equation 5 is that $\mathbf{h}_i$ can be seen as a baseline. The function $I_i(\mathbf{u}, 0)$ measures the likelihood that folds fans away from the singularity used to generate $\mathbf{h}_i$, as shown in Fig. 3(a), while the function $I_i(\mathbf{u}, \pi/2)$ measures the likelihood that folds rotate around the same singularities, as shown in Fig. 3(b).

The streamline tracing algorithm works as follows. Once it is determined that $\mathbf{h}_i$ is a locally optimal hypothesis an initial streamline is launched with an angle of α with $\mathbf{h}_i$ (for example $\alpha = \pi/2$). We then follow the flow given by the function $\mathbf{h}_i + \alpha$ over the entire region for which $\mathbf{h}_i$ is the optimal hypothesis. Examples of streamlines generated by this process are shown in Figure 3(c).



(a) Vector field in the vicinity of a sink

(b) Vector field curling around a singularity

(c) Streamlines inferred using curvatures and initial angles from Fig.3(b).

Fig. 3.

2.3 Relaxation Labeling of Tangential and Normal Curvature

Till now we have described an algorithm to generate flow fields $\mathbf{h}_i$ on a surface $\mathcal{S}$. In this section, we describe how to select, from $\mathbf{h}_i$, an optimal assignment that best matches the flow field oberved on a surface. Let $H^*(\mathbf{u}, \theta) \in \{\mathbf{h}_1, \ldots, \mathbf{h}_n\}$. We use relaxation labeling [8] to determine which of the possible flows $\mathbf{h}_i$ offers the best local fit to the fold lines of the cortical surface. Relaxation labeling is a framework to find the statistical mode of a distribution given general constraints to be satisfied. Let $p_i(\mathcal{S}, \theta)$ be the probability that hypothesis $\mathbf{h}_i(\mathcal{S}, \theta)$ has the highest support at any given location. Here, we interpret $\mathbf{h}_i(\mathcal{S}, \theta)$ has the hypothesis given by $\mathbf{h}_i$ with a flow field with initial angle θ. The $p_i(\mathcal{S}, \theta)$ forms a probability space, such that $p_i(\mathcal{S}, \theta) > 0$ and $\sum_{i=1}^{N} p_i(\mathcal{S}, \theta) = 1$ everywhere.

Relaxation labeling updates the p_i in order to maximize the local fit while making sure that each label is well supported by its neighbors. In our case, the optimal solution should (1) maximize the function f in Eq. 3 and (2) find solutions that are supported over a large spatial region.

To determine the solutions that have a large support on the manifold, we minimize the gradient of p_i along both $\mathcal{S}$ and θ as follows. Let $\mathbf{p}'_i$ be the resampled value of $\mathbf{p}_i$ along $\mathbf{h}_i$ as

$$\mathbf{p}'_i(\mathbf{u}, \theta) = \mathbf{p}_i(\mathbf{u}, \theta + \mathbf{h}_i) \ . \tag{7}$$

Then, we regularize the assignment $\mathbf{p}_i$ by minimizing the spatial and orientation gradient of $\mathbf{p}'_i$ as

$$U_{reg}(p_i) = (\|\nabla_{\mathcal{S}} p'_i\|^2) \circ \mathbf{h}_i^{-1} + \lambda_1 \|\nabla_\theta p_i\|^2 \ , \tag{8}$$

where $\circ \mathbf{h}_i^{-1}$ is used as a short-hand to mean that we should resample $(\|\nabla_{\mathcal{S}} p'_i\|^2)$ along the original θ orientation.

A relaxation labeling framework is then used to maximize the following functional

$$\mathbf{p}^* = argmax_{\mathbf{p}} \sum_i \int_{\mathcal{S}\times\mathbb{S}} -U_{reg}(p_i) + \lambda_2 p_i f_i + \lambda_3 \|p_i\|^2 d\mathcal{S} \wedge d\theta \tag{9}$$

where the $\|p_i\|^2$ term is added to make the relaxation labeling scheme converge to an unambiguous labeling, i.e., $p_i = \{0, 1\}$, as explained in [8].

2.4 Statistical Tests over Curve Length

We now estimate how curve length is affected by the presence of external factors. First, we launch streamline flows in multiple directions at every location on the surface. In practice we are not given trajectories, but rather an optimal assignment H^* given by the probabilities p_i. Let $\gamma(l)$ be a curve parametrized by arc-length and $\mathbf{t}$ the tangent of $\gamma(l)$. The curvature of $\gamma(l)$ (and hence its trajectory) is determined from the assignment p_i as

$$\frac{\partial}{\partial l}\mathbf{t} = \sum_i p_i(\gamma, \mathbf{t}) \left[\frac{\partial}{\partial l}\mathbf{h}_i\right] \ . \tag{10}$$

In practice, we use a first order Euler method to trace streamlines on surfaces. We take a small step in a given direction, then we compute the change in angle that is prescribed by Equation 10 given the current position and the given angle. Curve length is then measured explicitly as

$$L(\gamma) = \int_{[-T,T]} \sqrt{\|\frac{\partial \gamma}{\partial s}\|^2} ds \ . \tag{11}$$

Let $L(\gamma(\mathbf{u}, \theta))$ be the length of the curve started at $\mathbf{u} \in \mathcal{S}$ at a given orientation θ. The reader may realize that we have previously defined $\|\mathbf{t}\| = 1$ and thus $L(\gamma) = 2T$. However, Equation 11 becomes useful when we use a shape diffeomorphism onto a template space. Assume that there are n surfaces $\mathcal{S}_l, l = 1, \ldots, n$ with a set of diffeomorphisms that map these surfaces onto a template average $\bar{\mathcal{S}}$:

$$\phi_l : \mathcal{S}_l \rightarrow \bar{\mathcal{S}}. \tag{12}$$

Then, a random field can be defined by measuring the length of the curve γ when mapped using ϕ_l onto $\mathcal{S}_l$ as $L(\phi_l(\gamma(\mathbf{u}, \theta)))$. We then assume that the logarithm of curve length follows a Gaussian distribution, thus allowing us to define a Gaussian random field on $\mathcal{S} \times \mathbb{S}$. Some regularization and random field theory is then used to correct for multiple comparisons (see [9]).

3 Results

To present result, we first illustrate the algorithm by generating curves using a flow field $\mathbf{h}_i$. We use a template surface and then generate a full estimate of the curvature of the texture manifold using Equation 10. Several streamlines are launched in the direction of sulcal lines, as shown in Fig. 4. Observe that these streamlines bend and fan to follow the folds, qualitatively demonstrating the accuracy of the recovered curvature field.

We used the OASIS database [7] to determine if our method could find curves whose length is significantly correlated with the presence of mild cognitive impairment in Alzheimer's disease (AD). The OASIS database consists of 97 healthy subjects and 92 subjects (aged 60 and above) affected with mild and very-mild dementia. We used the extraction pipeline of [3], which produces one mid-surface representation of the gray-matter cortical sheet, and obtained mappings ϕ_l for each surface onto a common template average. Once this mapping was found, we computed the average surface of the entire population and used this average surface to compute curvature estimates using Equation 10.

The lengths of the computed curves were then tested to see if any significant correlations could be found with the presence of mild cognitive impairment in the OASIS database. The results are shown in Figure 6. The white streamlines indicate a significant dilation while the green streamlines show a significant contraction. These results show that the main shape differences are both perpendicular and parallel to the fold direction. The contraction is in the temporal

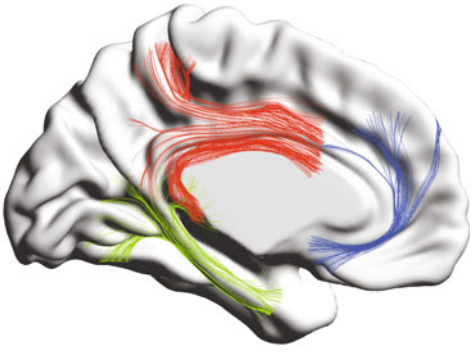


Fig. 4. Examples of streamlines of length 8 cm launched in directions that are tangent to sulcal lines. The streamlines bend and fan to follow the fold patterns. There are multiple gyrii fanning away in the vicinity of the blue lines.

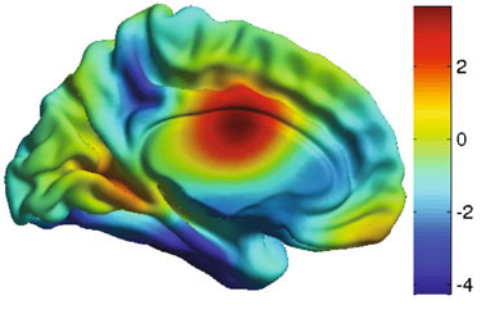
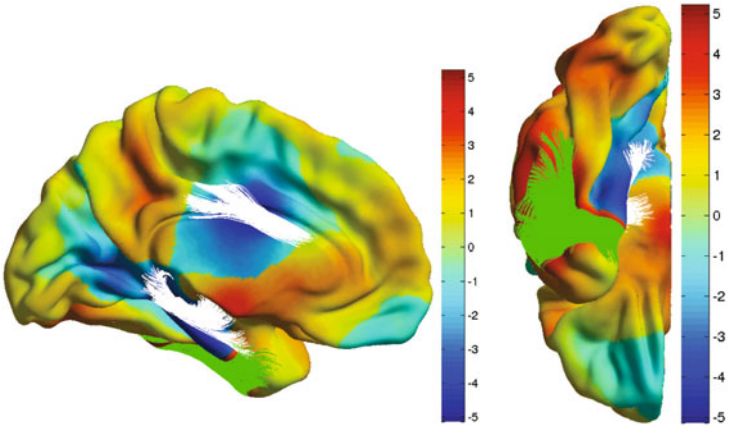


Fig. 5. Comparisons of the surface area of patients with mild cognitive impairment (MCI) and healthy subjects. A statistical t-Test (indicated by the colorbar) shows that the region below the Cingulate cortex has a larger surface area in the group with MCI. However, the t-Map does not reveal whether the larger area comes from a wider or a longer sulcus.

lobe, which is known to be affected with the presence of Alzheimer's disease. The dilation in the temporal lobe is probably due to atrophy of the hippocampus.

The most significant result is that our method is able to identify a stretching in length of the Cingulate gyrus below Praecuneus area and the area below the sup frontal cortex in the left hemisphere. We also performed a statistical test on the surface area [10] and these results are shown in Fig. 5. Whereas this test detects the same region (below the Cingulate gyrus) it does not reveal if the larger area comes from a wider or a longer sulcus.

Another interesting aspect of our method is that it integrates deformation along a very narrow streamline, permitting the use of an anisotropic kernel for regularization. Thus, the results using surface area were not significant in the region below the Cingulate cortex after correction for multiple comparison using Random Field correction [9]. However, the use of anisotropic filters produced



(a) Saggital cut of the left hemisphere. The t-Test shows that the larger area in Fig. 5 is due to a longer Cingulate gyrus in the group with MCI.

(b) Bottom view

Fig. 6. Positive values shows significant dilation in the AD group. White streamlines show a significant dilation of the curve length ($p < 0.0015$ on top of the hyppocampal gyrus and $p < 0.01$ on the cingulate gyrus) and green streamlines in the temporal lobe show a significant contraction ($p < 0.0007$). Color indicates the maximum in absolute value of the t-test over all possible orientations.

significant results in both the temporal lobe and the Cingulate cortex, as shown in Fig. 6. The threshold for significance are $|t| > 4.25$ using Random Field Theory and $|t| > 4.31$ using a permutation test with 50000 permutations. The peak values of the three regions shown in Fig. 6 were above the permutation test threshold and the p values reported in the caption of Figure 6 are computed using Random Field correction [9].

Overall these results show that it is possible using directional filters to gain an insight into the process that leads to brain deformations. Thus, wether the stretch in length is due to structural connectivity loss in the white matter still needs to be investigated, but our findings are corroborated by a visual inspection of shape differences.

4 Conclusion

We have described a method to perform a dense statistical analysis of the cerebral cortex using curve-based morphometry. Our method departs from other statistical methods, e.g., those based on shape diffeomorphisms. We define a manifold of curves on the cerebral cortex and then use statistics on curve length to examine shape changes. We obtain novel results concerning the nature of changes in cortical fold patterns in subjects with mild cognitive impairment.

References

1. Fischl, B., Sereno, M.I., Dale, A.M.: Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage 9(2), 195–207 (1999)
2. Leporé, N., Brun, C., Pennec, X., Chou, Y., Lopez, O., Aizenstein, H., Becker, J., Toga, A., Thompson, P.: Mean template for tensor-based morphometry using deformation tensors. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 826–833. Springer, Heidelberg (2007)
3. Lyttelton, O., Boucher, M., Robbins, S., Evans, A.: An unbiased iterative group registration template for cortical surface analysis. Neuroimage 34(4), 1535–1544 (2007)
4. Worsley, K.: Testing for signals with unknown location and scale in a chiˆ 2 random field, with an application to fMRI. Advances in Applied Probability 33(4), 773–793 (2001)
5. Chung, M., Worsley, K., Evans, A.: Tensor-based brain surface modeling and analysis. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 467–473 (2003)
6. Ben-Shahar, O., Zucker, S.: The perceptual organization of texture flow: A contextual inference approach. PAMI 25(4), 401–417 (2003)
7. Marcus, D., Wang, T., Parker, J., Csernansky, J., Morris, J., Buckner, R.: Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults. Journal of Cognitive Neuroscience 19(9), 1498–1507 (2007)
8. Hummel, R., Zucker, S.: On the foundations of relaxation labeling processes. In: Readings in computer vision: issues, problems, principles, and paradigms, p. 605. Morgan Kaufmann Publishers Inc., San Fransico (1987)
9. Adler, R., Taylor, J.: Random fields and geometry. Springer, New York (2007)
10. Chung, M., Worsley, K., Robbins, S., Paus, T., Taylor, J., Giedd, J., Rapoport, J., Evans, A.: Deformation-based surface morphometry applied to gray matter deformation. NeuroImage 18(2), 198–213 (2003)

Semisupervised Probabilistic Clustering of Brain MR Images Including Prior Clinical Information

Annemie Ribbens, Frederik Maes, Dirk Vandermeulen, and Paul Suetens

Katholieke Universiteit Leuven, Faculty of Engineering, Department of Electrical Engineering - ESAT, Center for Processing Speech and Images - PSI
annemie.ribbens@uz.kuleuven.ac.be

Abstract. Accurate morphologic clustering of subjects and detection of population specific differences in brain MR images, due to e.g. neurological diseases, is of great interest in medical image analysis. In previous work, we proposed a probabilistic framework for unsupervised image clustering that allows exposing cluster specific morphological differences in each image. In this paper, we extend this framework to also accommodate semisupervised clustering approaches which provides the possibility of including prior knowledge about cluster memberships, group-level morphological differences and clinical prior knowledge. The method is validated on three different data sets and a comparative study between the supervised, semisupervised and unsupervised methods is performed. We show that the use of a limited amount of prior knowledge about cluster memberships can contribute to a better clustering performance in certain applications, while on the other hand the semisupervised clustering is quite robust to incorrect prior clustering knowledge.

1 Introduction

Many neurological diseases involve macroscopic effects visible in brain MR images, in particular morphological changes of anatomical structures. Identification of such disease related morphological features contributes to the clustering of the images according to disease, which is important for early diagnosis. Clustering of images based on spatial features and detection of relevant features are challenging tasks for which different algorithms have been proposed [1,2,3,4,5,6,7].

Most approaches are supervised requiring a training set of clustered images to construct clustering rules (using support vector machines for instance) based on group-level differences defined in terms of segmentations [1], deformations [6] and/or other features [2]. Unsupervised methods [5,7] on the other hand have the advantage that no prior knowledge about the cluster memberships of any of the images is needed. Therefore, they can contribute to the discovery of subgroups which differ in unexpected ways. Moreover, the clustering rule (and the identification of disease related effects that is based on it) is not affected by possibly incorrectly clustered images in the training set.

Combining both approaches leads to semisupervised clustering, i.e. incorporating available prior knowledge about the cluster memberships of some of the

B. Menze et al. (Eds.): MICCAI 2010 Workshop MCV, LNCS 6533, pp. 184–194, 2011.

images to improve the clustering of new uncategorized images. On the other hand, the discovery of new subpopulations that differ in unexpected ways remains possible, but can be penalized. Moreover, we expect such a framework to be more robust to incorrectly clustered training images than a completely supervised algorithm. Multiple semisupervised learning and clustering approaches are proposed in literature [8,9]. In this paper, we further explore the unsupervised clustering framework SPARC [7] and extend it to a semisupervised clustering algorithm as it is already proven that the unsupervised framework gives good results when clustering brain MR images according to disease. Furthermore, the framework also delivers automatically the image segmentations, the mean template per cluster and indicates the disease-specific morphological differences. In the next section, the methods are described. The third section compares semisupervised SPARC with its supervised and unsupervised variants.

2 Methods

2.1 Unsupervised SPARC

SPARC is a unified probabilistic framework that, given a set of brain MR images of different subjects and a pre-set number of clusters, iteratively and simultaneously **S**egments the images in tissue classes, constructs **P**robabilistic **A**tlases per cluster, performs nonrigid atlas-to-image **R**egistration and estimates the fuzzy **C**luster memberships for each image. The cluster memberships are defined voxelwise, such that SPARC can identify cluster-specific spatially localized features and has the potential to expose specific morphological differences between population subgroups. Denote $Y = \{y_{ij}\}$ the image intensities, $L = \{l_{ijk}\}$ the tissue segmentations and $Z = \{z_{ijt}\}$ the cluster memberships with $i \in \{1, \ldots, N_I\}$, $j \in \{1, \ldots, N_J\}$, $t \in \{1, \ldots, N_T\}$ and $k \in \{1, \ldots, N_K\}$ indexing images, voxels, clusters and tissue classes respectively and N_T and N_K are provided by the user. The model assumes per image i a Gaussian for each tissue class k (with Gaussian parameters $\theta_{ik} = \{\mu_{ik}, \sigma_{ik}^2\}$, i.e. mean and variance) on the bias field corrected image intensities (bias field parameters C_i). This probability is denoted as $P(y_{ij}|\theta_{ik}, C_i)$. Furthermore, a cluster-specific prior is defined on the tissue labels $P(l_{ijk}|A_{kt}, R_{ijt})$ as the cluster atlas A_{kt} after nonrigid registration to each image through the deformation fields R_{ijt}. A Gaussian prior $P(R_{ijt}|G_{jt}, \epsilon_{jt}^2)$ per cluster with mean G_{jt} and variance ϵ_{jt}^2 is put on the deformations to restrict these drifting away too far from the groupwise deformation G_{jt}. Finally, a prior is added on the cluster memberships of each voxel $P(z_{ijt}|\pi_{it})$. For unsupervised clustering, no prior clustering information is assumed and a uniform distribution, equal for all voxels of the same image, is used.

The model is then formulated as a maximum a posteriori (MAP) problem and optimized using an iterative expectation maximization (EM) procedure with observed variables the image intensities Y, hidden variables the cluster memberships Z and tissue class labels (segmentations) L and model parameters $\Upsilon = \{\theta_{ik}, C_i, A_{kt}, R_{ijt}, G_{jt}, \epsilon_{jt}^2 | \forall i, j, k, t\}$. In the E-step of the EM algorithm

the expectation $Q(\Upsilon|\Upsilon^\eta)$ of the log-likelihood function $\log P(Y|\Upsilon^\eta)$ given the current estimate Υ^η of the model parameters Υ is determined:

$$Q(\Upsilon|\Upsilon^\eta) \propto \sum_{i=1}^{N_I}\sum_{j=1}^{N_J}\sum_{k=1}^{N_K}\sum_{t=1}^{N_T} b_{ijkt}^{(\eta+1)} \cdot [\log P(y_{ij}|\theta_{ik}, C_i) + \log P(l_{ijk}|A_{kt}, R_{ijt}) + \log P(R_{ijt}|G_{jt}, \epsilon_{jt}^2) + \log P(z_{ijt}|\pi_{it})] \quad (1)$$

with η the iteration number and $b_{ijkt}^{(\eta+1)} = P(l_{ijk}, z_{ijt}|y_{ij}, \Upsilon_{ijkt}^\eta)$ the posterior distribution of the hidden variables determined using Bayes' rule. In the M-step the parameters Υ are updated by maximization of Q, except ϵ_{jt}^2 which is chosen in advance. All solutions are closed form, except for the atlas-to-image registration. Therefore and to obtain a physically acceptable deformation field, the viscous fluid regularizer [10] is used. The EM iterations are stopped when the increase of the log-likelihood $P(Y|\Upsilon^\eta)$ becomes insignificant. A detailed description of unsupervised SPARC and its parameters can be found in [7].

2.2 Semisupervised SPARC

Weak supervision in clustering methods formulated as an EM algorithm can be incorporated in several ways. Often extra prior information about the cluster memberships of some data is given in the form of must-link and cannot-link constraints (e.g. [11]). Here, we incorporate this type of information in the SPARC framework by imposing a Markov Random field (MRF) on the cluster membership prior, instead of using a uniform prior as in equation (1) above. Such MRF prior is defined by the Gibbs distribution:

$$P(z_{ijt}|z_{\mathcal{N}_{ij}}, \Phi_{z_{ijt}}) = Z(\Phi_{z_{ij}})^{-1} \exp\left[-U(z_{ijt}|z_{\mathcal{N}_{ij}}, \Phi_{z_{ijt}})\right] \quad (2)$$

with $\Phi_{z_{ijt}}$ the MRF parameters, $z_{\mathcal{N}_{ij}}$ the cluster memberships of voxels in a neighborhood $\mathcal{N}_{ij}$ of voxel j of image i and $Z(\Phi_{z_{ij}}) = \sum_t \exp\left[-U(z_{ijt}|z_{\mathcal{N}_{ij}}, \Phi_{z_{ijt}})\right]$. The function U is an energy function with parameters Φ that is defined as the sum of clique potentials over the neighborhood $\mathcal{N}_{ij}$, as described below. The calculation of this prior and of the log-likelihood itself requires all possible realizations of the MRF, which is computationally not feasible. Therefore, we use the mean field approximation [12,13] to derive the E-step:

$$b_{ijkt}^{(\eta+1)} \approx \frac{P(y_{ij}|\theta_{ik}^\eta, C_i^\eta) P(l_{ijk}|A_{kt}^\eta, R_{ijt}^\eta) P(z_{ijt}|z_{\mathcal{N}_{ij}}^\eta, \Phi_{z_{ijt}}) P(R_{ijt}|G_{ijt}^\eta, \epsilon_{jt}^{2\eta})}{\sum_k \sum_t P(y_{ij}|\theta_{ik}^\eta, C_i^\eta) P(l_{ijk}|A_{kt}^\eta, R_{ijt}^\eta) f(z_{ijt}|z_{\mathcal{N}_{ij}}^\eta, \Phi_{z_{ijt}}) P(R_{ijt}^\eta|G_{ijt}^\eta, \epsilon_{jt}^{2\eta})}$$

while the M-step remains similar to unsupervised SPARC [7].

The energy function U of the random field can be generally defined by the following Potts model:

$$U(z_{ijt}|z_{\mathcal{N}_{ij}}, \Phi_{z_{ijt}}) = \sum_{i',j' \in \mathcal{N}_{ij}} \sum_{t'} z_{ijt} \alpha_{ii'jj'tt'} z_{i'j't'} \quad (3)$$

with $\mathcal{N}_{ij}$ the neighborhood of voxel j in image i and $\Phi_{z_{ijt}} = \{\alpha_{ii'jj'tt'} | i', j' \in \mathcal{N}_{ij}, \forall t'\}$ the MRF parameters. The neighborhoods and MRF parameters are defined based on the available prior knowledge. If prior knowledge about cluster memberships of some images is available, these images are defined to belong to the same neighborhood. Both must-link and cannot-link constraints can be incorporated by choosing the MRF parameters such that they penalize images with different cluster memberships and stimulate images with equal cluster memberships to belong to the same cluster. If the use of only must-link constraints is preferred, different neighborhoods for different clusters can be defined.

In SPARC, cluster memberships are defined at the voxel-level, each contributing to the clustering of the entire image, such that constraints between images as well as between voxels (in the same or in different images) can be incorporated. Voxelwise constraints allow to incorporate spatial information. Neighboring voxels or voxels corresponding to the same anatomical structure can be forced to belong to the same cluster based on neighborhood restriction or MRF parameters. Practically this means that identification of specific group-level morphological differences becomes region based. Moreover, voxelwise constraints allow incorporating prior knowledge about the group-level differences. For instance, if it is known or hypothesized that a particular anatomical structure is significantly different between two populations, the contribution of voxels in a region of interest could be amplified for image clustering. This can be achieved by setting the parameter $\alpha_{ii'jj'tt'}$ smaller in case $t = t'$ and larger for $t \neq t'$ if images i and i' are stimulated to belong to the same cluster and vice versa if images i and i' are penalized to belong to the same cluster. For the experiments in this paper, we concentrate on must-link constraints between images without discriminating between voxels by averaging the voxelwise cluster memberships per image.

2.3 Supervised SPARC

Supervised SPARC is performed as analogously as possible to standard supervised clustering algorithms, constructing a clustering rule based on a training set. The same clustering rule as in unsupervised and semisupervised SPARC is used, i.e. the sum over all tissue classes of the posterior b_{ijkt}

$$\text{decision rule} = \frac{\sum_k P(y_{ij}|\theta_{ik}, C_i) P(l_{ijk}|A_{kt}, R_{ijt}) P(z_{ijt}|\pi_{it}) P(R_{ijt}|G_{jt}, \epsilon_{jt}^2)}{\sum_t \sum_k P(y_{ij}|\theta_{ik}, C_i) P(l_{ijk}|A_{kt}, R_{ijt}) P(z_{ijt}|\pi_{it}) P(R_{ijt}|G_{jt}, \epsilon_{jt}^2)}$$

where the atlases A_{kt} and groupwise registrations G_{jt} are determined based on the training subjects per cluster using SPARC with one cluster. The prior on the clustering π_{it} is fixed to 0.5. When a new subject needs to be classified, the Gaussian mixture model and the atlas-to-image registration for each cluster are obtained using SPARC (one cluster) without updating the atlas and the parameter G_{jt}, which is similar to the framework described in [14].

3 Experiments and Results

We focus in the experiments on validating the performance of unsupervised versus semisupervised clustering strategies embedded in the SPARC framework, although it is to be expected that improved clustering also leads to more specific atlases and improved segmentations.

A first experiment is performed on 30 skull stripped and bias field corrected images of the publicly available OASIS data set [15], the same as used in [5]. This subset consists of 15 images of patients diagnosed with mild dementia and probable Alzheimer disease (AD) and 15 images of normal aged persons. For each subject, the score of the mini-mental state examination (MMSE) gives a clinical indication to which cluster (normal aged or demented) it belongs. First, we apply unsupervised SPARC to classify the images into two clusters without prior clustering information, i.e. using a uniform prior for the cluster memberships. Secondly, we include the MMSE scores as prior information in semisupervised SPARC, defining the MRF parameters to incorporate 'soft' must-link and cannot-link constraints based on differences in MMSE scores:

$$\alpha_{ii'jj'tt'} = \begin{cases} \beta_{i'j'} \cdot |\text{rMMSE(i)} - \text{rMMSE(i')}| & t = t' \\ \beta_{i'j'} \cdot [1 - |\text{rMMSE(i)} - \text{rMMSE(i')}|] & t \neq t' \end{cases}$$

where rMMSE represents a rescaled MMSE score determined based on a regression analysis using all data in the OASIS data set and $\beta_{i'j'}$ is a weighting factor that favors voxels of the same image $i = i'$ over the ones of neighboring images to determine the cluster memberships. The MRF parameters can be multiplied with a constant, which can be interpreted as the inverse of the temperature. Starting with a high initial value for the temperature, while decreasing it during iterations, can avoid to end up in local optima. Finally, clustering was also performed using the clinical information only by deriving the fuzzy cluster memberships from regression analysis of the MMSE scores directly.

Table 1(a) lists the obtained fuzzy cluster memberships for the three clustering schemes (i.e. using clinical information only, unsupervised SPARC based on morphological information only, and semisupervised SPARC using both clinical and morphological information). Semisupervised SPARC classifies 30/30 images correctly, versus 29/30 for unsupervised SPARC. Both results compare favorably against [5], where 10/15 images of the demented group were misclassified as normal. To illustrate the disease related morphological differences, Figure 1 shows the difference between both obtained atlases (normal aged and demented) and the cluster-specific spatial features for a particular normal aged and a particular demented subject as exposed with unsupervised SPARC. We remark a more extensive overall shrinkage of brain tissue in brains of mild dementia and Alzheimer's disease patients compared to normal aged brains, i.e. the sulci are more widened, the gyri more shrunken and the ventricles more enlarged. We also observe a little more volume loss in the hippocampal region. This is in correspondence with what is found in literature [16].

A second experiment is performed on a data set containing 15 3D T1-weighted MR images of healthy young persons (20 - 25 year) and 15 images of healthy elderly (66 - 73 year). All images have image dimensions of $256 \times 256 \times 182$ and voxel sizes around 1mm^3 and are affinely aligned to the SPM-MNI space. To assess the impact of including prior knowledge on the clustering performance, we apply both unsupervised SPARC and semisupervised SPARC for clustering the images according to age group. In semisupervised SPARC, we assume cluster memberships are given for 4 images in each cluster (about 25% of the images). Must-link constraints are imposed on these images per cluster wherefore a special neighborhood per cluster is created containing the given 4 images. The MRF parameters determining the strength of the must-link interactions between the different images should be defined such that images that are more similar are stimulated more to belong to the same cluster. Here this similarity measure is chosen to be based on mutual information between the affinely registered images, but other (e.g. segmentation based) measures are equally feasible.

Table 1(b) lists the fuzzy cluster memberships obtained using unsupervised and semisupervised SPARC. The unsupervised clustering performance is rather poor here: it has not really diverged in two groups and 6 images are classified incorrectly. In semisupervised SPARC, only 3 images are classified incorrectly and the clustering is more diverged in two groups. This confirms the results of the first experiment, i.e. incorporation of prior clustering knowledge can improve the clustering performance.

A third experiment is performed on a data set containing 7 3D T1-weighted MR images of healthy persons and 7 images of Huntington disease patients (HD). All images have image dimensions of $256 \times 256 \times 182$, voxel sizes around 1mm^3 and are affinely aligned to the SPM-MNI space. We assess the robustness of supervised and semisupervised SPARC for assigning an image to the correct cluster when a significant number of the training data have an incorrect cluster label. For a fair comparison, all available images, except the image to be classified, are used as training data in both methods (leave-one-out). For semisupervised SPARC, this means that must-link constraints are imposed on all images per cluster, except the image to be classified. The MRF parameters are again based on mutual information between the images and the cluster memberships of all voxels of all images in a neighborhood contribute equally.

Table 1(c) lists the fuzzy cluster memberships obtained using SPARC with different levels of supervision and a training set containing 30% incorrectly labeled images. Supervised SPARC can not correct for the misclassified training images, leading to an incorrect classification of the new subject. Semisupervised SPARC assigns the new subject to the correct cluster, which demonstrates that it can successfully cope with misclassified training data. Three of the four misclassified training images are assigned a fuzzy cluster membership of exactly 0.5, while also all other values are near 0.5. This indicates that the training set contains misclassified images as otherwise the must-link constraints would stimulate divergence of the fuzzy cluster memberships. For unsupervised SPARC, no training set was needed. Accurate cluster-specific atlases are constructed, revealing

Table 1. Fuzzy cluster memberships (rescaled to 1000) for the three experiments for supervised (sup), semisupervised (semi) and unsupervised (un) SPARC. Green = correct, Red = incorrect, Orange = undecided cluster memberships.

(a) Experiment 1: Fuzzy cluster memberships to belong to the cluster of normal aged brains. The fuzzy cluster memberships rMMSE are obtained using a regression analysis on the clinical MMSE score.

Normal aged															
Subject	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Un	504	503	507	510	500	494	502	504	508	505	505	505	511	503	505
Semi	547	537	553	566	513	501	520	543	564	547	547	509	576	541	544
rMMSE	842	842	842	1000	708	708	708	842	1000	842	842	495	1000	842	842
Demented (AD)															
Subject	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
Un	498	499	494	488	497	488	496	491	498	489	489	490	487	491	488
Semi	456	499	448	428	412	454	426	446	438	464	426	439	425	442	423
rMMSE	0	593	174	221	134	19	134	0	221	275	133	275	42	221	0

(b) Experiment 2: Fuzzy cluster memberships to belong to the elderly cluster. Images with must-link constraints in semisupervised SPARC are underlined.

Elderly															
Subject	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Age	66	66	63	63	66	67	62	62	63	66	65	70	70	69	73
Un	515	532	525	509	507	484	509	478	519	506	511	506	500	511	509
Semi	516	517	516	504	460	490	560	515	540	500	517	555	507	578	523
Young															
Subject	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
Age	22	23	23	20	20	22	21	22	22	24	23	23	23	25	22
Un	499	505	489	493	484	496	493	506	505	493	502	500	479	490	469
Semi	483	478	474	478	479	517	422	470	480	461	445	480	479	421	480

(c) Experiment 3: Fuzzy cluster memberships to belong to the cluster of normals. The cluster memberships determined by the clinician (real) and the initialization (init) including the misclassified data are given. Image 8 is the image to be classified.

	Normal							**HD**						
Subject	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Real	1000	1000	1000	1000	1000	1000	1000	0	0	0	0	0	0	0
Init	1000	1000	1000	1000	0	1000	0	?	0	1000	1000	0	0	0
Sup	1000	1000	1000	1000	0	1000	0	505	0	1000	1000	0	0	0
Semi	504	507	506	505	500	505	500	495	494	505	500	493	489	491
Un	648	607	475	549	654	579	648	407	500	358	317	468	368	295

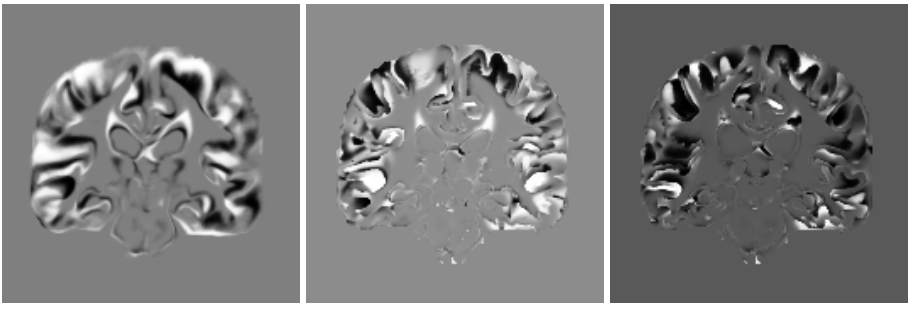


Fig. 1. Experiment 1: Results obtained using unsupervised SPARC with two clusters (AD and normal). Left: Difference image between gray matter maps of the normal and AD atlas. Middle and Right: Cluster-specific spatial features for a particular normal aged subject (middle) and a particular AD subject (right), i.e. voxelwise cluster memberships to belong to the cluster of normal aged brains (brigher/darker means more likely to belong to the normal/AD cluster).

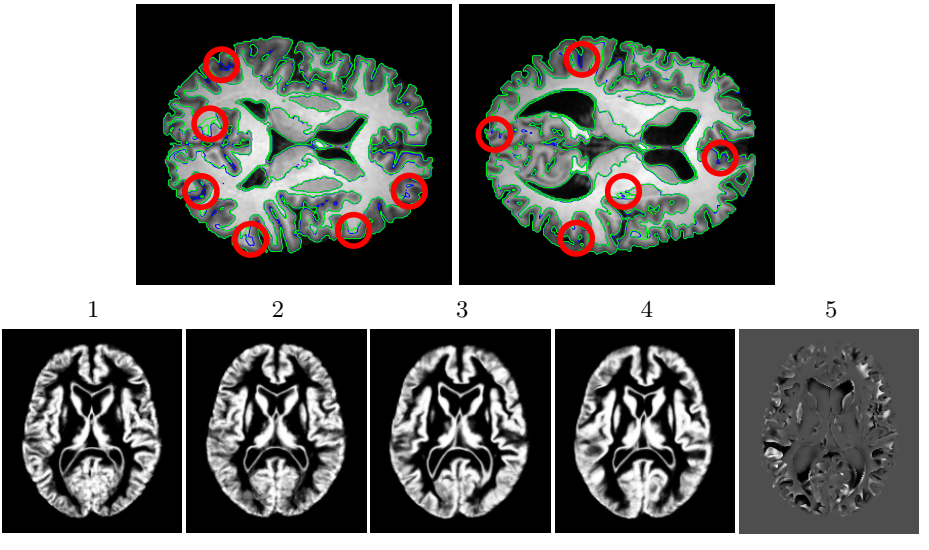


Fig. 2. Experiment 3: Top: Gray matter segmentations for two images using unsupervised (blue) and supervised (green) SPARC. Red circles indicate sites where unsupervised SPARC outperforms supervised SPARC. Bottom: (1,2) Atlases for HD and normal clusters obtained using supervised SPARC with misclassified images in the training set, (3,4) idem using unsupervised SPARC, (5) cluster-specific spatial features for a particular HD subject as exposed with unsupervised SPARC (i.e. voxelwise cluster memberships, brighter/darker means more likely to belong to normal/HD cluster).

morphological disease-specific differences between both clusters (as illustrated in Figure 2). Unsupervised SPARC clusters almost all images correctly. The two incorrectly clustered images were visually inspected (ventricles, deep gray matter structures) and were found to be on the border of HD and normal. A final remark is that the segmentation quality of unsupervised SPARC seems to outperform that of supervised SPARC, even when a training set without misclassified images is used (Figure 2). This sounds contradictory to the fact that we expect a better segmentation performance if a better clustering performance is obtained. However, the better segmentation with unsupervised SPARC is observed especially in regions where large inter-subject variability occurs, but no inter-cluster variability. Hence, more images contribute to the segmentation of these regions, as the cluster memberships are determined voxelwise.

4 Discussion and Conclusion

SPARC is a unified probabilistic framework that simultaneously segments a set of images in tissue classes and clusters them into different subpopulations without the need for prior knowledge. The method automatically generates non-rigid probabilistic atlases for each subpopulation, which can serve as specific atlases for more basic segmentation algorithms. It also reveals the localization of cluster-specific morphological differences for each image. The unified framework of SPARC makes that all these aspects can benefit from each other. We applied the framework here to three different data sets and showed that SPARC is able to cluster a set of images into different subpopulations and find the cluster related morphological differences. We extended SPARC to a semisupervised algorithm by including must-link and cannot-link constraints using a MRF on the cluster memberships. This allows incorporating different types of clinical prior knowledge during clustering, while the fact that the MRF and cluster memberships are defined at the voxel-level introduces the possibility of easily including region-specific prior information and prior knowledge about group-level differences.

We compared the semisupervised framework to its supervised and unsupervised variants. In the first two experiments, it was demonstrated that incorporating prior knowledge of different types (such as mental stage, age, known cluster memberships, etc.) can contribute to a better clustering performance. The advantages of semisupervised over supervised clustering are that less prior knowledge is needed, that the potential of discovering new subpopulations is not lost and that it is less sensitive to incorrect prior clustering information, (e.g. a training set with incorrectly classified images), as was illustrated in the third experiment. This experiment also showed that unsupervised SPARC (and also semisupervised SPARC) can yield locally better segmentation quality than supervised SPARC. This can be explained by the fact that unsupervised and semisupervised SPARC combine the information of all images in all clusters and systematically select for each region the more appropriate cluster-specific atlas for the segmentation of the images, which has been shown to improve segmentation performance [17].

The few misclassifications in our experiments with unsupervised and semisupervised SPARC may be attributed to the subtility of the morphological changes induced by neurological disease or degeneration, by the fact that relevant spatial features are not so clear in some data sets or appear in regions where the intragroup variability is already quite large, and by our choice of parameters (e.g. ϵ^2 in the prior on the deformation). Future work will focus on a more extensive validation with specific attention to parameter settings such as the choice of the MRF parameters, in particular suitable image-based measures for defining clustering affinities. Further, a study of the influence of the size of the data set, of the number of training data and of incorporating region based prior knowledge is of interest, as for instance larger data sets lead to a better averaging of intraclass variability which can contribute to the clustering performance. Finally, it is worthwhile to compare our algorithm with other semisupervised clustering methods proposed in literature.

References

1. Ashburner, J., Friston, K.J.: Voxel-based morphometry - the methods. NeuroImage 11, 805–821 (2000)
2. Liu, Y., Teverovskiy, L., Carmichael, O., Kikinis, R., Shenton, M., Carter, C.S., Stenger, V.A., Davis, S., Aizenstein, H., Becker, J., Lopez, O., Meltzer, C.: Discriminative MR image feature analysis for automatic schizophrenia and Alzheimer's disease classification. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3216, pp. 393–401. Springer, Heidelberg (2004)
3. Fan, Y., Shen, D., Davatzikos, C.: Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 1–8. Springer, Heidelberg (2005)
4. Duchesne, S., Caroli, A., Geroldi, C., Barillot, C., Frisoni, G.B., Collins, D.L.: MRI-based automated computer classification of probable AD versus normal controls. IEEE Trans. on Med. Img. 27, 509–520 (2008)
5. Sabuncu, M.R., Balci, S.K., Shenton, M.E., Golland, P.: Image-driven population analysis through mixture modeling. IEEE Trans. on Med. Img. 28, 1473–1487 (2009)
6. Pohl, K.M., Sabuncu, M.R.: A unified framework for MR based disease classification. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 300–313. Springer, Heidelberg (2009)
7. Ribbens, A., Hermans, J., Maes, F., Vandermeulen, D., Suetens, P.: SPARC: Unified framework for automatic segmentation, probabilistic atlas construction, registration and clustering of brain MR images. In: IEEE ISBI, pp. 856–859 (2010)
8. Grira, N., Crucianu, M., Boujemaa, N.: Unsupervised and semisupervised clustering: a brief survey. A Review of Machine Learning Techniques for Processing Multimedia Content, Report of the MUSCLE European Network of Excellence (2004)
9. Zhu, X.: Semi-supervised learning literature survey. Technical report, University of wisconsin at Madison (2006)
10. D'Agostino, E., Maes, F., Vandermeulen, D., Suetens, P.: A viscous fluid model for multimodal non-rigid image registration using mutual information. MedIA 7(4), 565–575 (2003)

11. Shental, N., Bar-Hillel, A., Hertz, T., Weinshall, D.: Computing Gaussian mixture models with EM using equivalence constraints. Neural Inf. Proc. Systems 16, 185–192 (2003)
12. Zhang, J.: The mean-field theory in EM procedures for Markov Random Fields. IEEE Trans. on Signal Processing 40, 2570–2583 (1992)
13. Langan, D., Molnar, K., Modestino, J., Zhang, J.: Use of the mean-field approximation in an EM-based approach to unsupervised stochastic model-based image segmentation. In: Proc. ICASSP, vol. 3, pp. 57–60 (1992)
14. D'Agostino, E., Maes, F., Vandermeulen, D., Suetens, P.: A unified framework for atlas based brain image segmentation and registration. In: Pluim, J.P.W., Likar, B., Gerritsen, F.A. (eds.) WBIR 2006. LNCS, vol. 4057, pp. 136–143. Springer, Heidelberg (2006)
15. Marcus, D., Wang, T., Parker, J., Csernansky, J., Morris, J., Buckner, R.: Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. Journal of Cognitive Neuroscience 19, 1498–1507 (2007)
16. Karasa, G., Scheltens, P., Romboutsc, S., Visserc, P., van Schijndel, R., Foxf, N., Barkhofa, F.: Global and local gray matter loss in mild cognitive impairment and Alzheimer's disease. NeuroImage 23, 708–716 (2004)
17. Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J., Rueckert, D.: Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. NeuroImage 46, 726–738 (2009)

Simultaneous Multi-object Segmentation Using Local Robust Statistics and Contour Interaction

Yi Gao[1], Allen Tannenbaum[1,2], and Ron Kikinis[3]

[1] Schools of Electrical Computer Engineering and Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, 30332
[2] Department of EE, Technion-IIT, Haifa 32000, Israel
[3] Surgical Planning Laboratory, Brigham & Women's Hospital, Harvard Medical School, Boston, MA 02115

Abstract. In this work, we present an active contour scheme to simultaneously extract multiple targets from MR and CT medical imagery. A number of previous active contour methods are capable of only extracting one object at a time. Therefore, when multiple objects are required, the segmentation process must be performed sequentially. Not only may this be tedious work, but moreover the relationship between the given objects is not addressed in a uniform framework, making the method prone to leakage and overlap among the individual segmentation results. On the other hand, many of the algorithms providing the capability to perform simultaneous multiple object segmentation, tacitly or explicitly assume that the union of the multiple regions equals the whole image domain. However, this is often invalid for many medical imaging tasks. In the present work, we give a straightforward methodology to alleviate these drawbacks as follows. First, local robust statistics are used to describe the object features, which are learned adaptively from user provided seeds. Second, several active contours evolve simultaneously with their interactions being governed by simple principles derived from mechanics. This not only guarantees mutual exclusiveness among the contours, but also no longer relies upon the assumption that the multiple objects fill the whole image domain. In doing so, the contours interact and converge to equilibrium at the desired positions of the given objects. The method naturally handles the issues of leakage and overlapping. Both qualitative and quantitative results are shown to highlight the algorithm's capability of extracting several targets as well as robustly preventing the leakage.

1 Introduction

Extracting anatomically and/or functionally significant regions from medical imagery, i.e., segmentation, is a challenge and important task in medical image analysis. One common practice consists of user initialization with one or several clicks (often called "seeds") in the target, and the algorithm then takes over to extract the desired object. A simple but intuitive example using such strategy is the region growing method [1]. Although the formalism is simple and straightforward, it reflects the two key roles of the user initialization: *Position*:

B. Menze et al. (Eds.): MICCAI 2010 Workshop MCV, LNCS 6533, pp. 195–203, 2011.

the positions of the initial seeds indicate the estimated position of the target; *Feature*: the image information in a given neighborhood of the seeds should be employed to learn the necessary characteristics of the desired object as well as to drive the segmentation. Nevertheless, original region growing only depends on the image intensity, and thus is many times not suitable for noisy and textured imagery. Furthermore, the segmentation boundary is not guaranteed to be as smooth as many times required. To address the first problem, Pichon *et al.* used robust statistics for better modeling of the image features at the locations of the seeds, and a fast marching algorithm to grow the segmentation contour [2]. Various active contour methods evolve a contour (curve or surface) in a variational manner to utilize both image information and contour geometry; see [3,4,5,6,7,8,9,10,11] and the references therein. The method proposed here follows this general philosophy, but in contrast to many active contour methods which only utilize the position information of the seeds, here we makes full use of the image information around the seeds in an adaptive fashion. Basically, the target object characteristics are learned online from the user inputs. Then the active contour evolves from the given places and converges to the desired boundary of the target.

Moreover, another desired feature for segmentation is the ability to simultaneously extract multiple objects. This can be quite advantageous in medical image analysis, where several related targets all need to be captured. However, most active contour algorithms are tailored to handle only one target at a time. Thus, the given algorithm needs to be executed sequentially several times in order to obtain the required multiple objects. However, since the individual segmentation processes do not interact with each other, it is difficult to guarantee mutual exclusiveness among contours. To address that, multiple object segmentation has been discussed in several papers [12,13,14,15,16,17]. In these works, the algorithms require the contours to be mutually exclusive (not overlapping). In addition, they also assume that the union of the regions bounded by the contours must be equal to the entire image domain. However, this is usually not a valid assumption for many medical imaging tasks. Our methodology does not rely on this assumption, which makes it more suitable for many medical imaging problems. This is accomplished by incorporating simple principles from mechanics into the contour interactions, which also handles the aforementioned problem of overlapping. Thus the algorithm naturally treats the issue of leakage. Moreover, researchers in [18,19] used the shape prior to achieve the multiple target objective. However, not only that requires the learning data set and process for the shape prior, but also the mutual exclusiveness among the contours are not guaranteed.

2 Method

If we consider the segmentation process in our own visual system, we observe that when human is recognizing the objects in a scene, several basic steps take place in sequence [20]. We will illustrate this via an example. Suppose that we want to trace out the boundary of both the liver and the right kidney in

medical imagery. First, prior anatomy knowledge drives our attention to the right abdominal region. Second, we focus at an area where we believe to be most "liver-like," and learn the liver characteristics in this particular image. With such knowledge, we then move our focus to enclose more tissue that looks similar to those representative regions. Usually, such similarity ends when we reach a remote area. In particular, at the boundary where the liver touches the right kidney, the decision is difficult. Under such a situation, we apply a similar procedure to the kidney, and we come back to the same ambiguous region. However, this time with the information from both sides (liver and kidney), internally we perform a competition: we compare the current voxel with both the liver and the kidney to decide which boundary should advance, so the other should retreat. Finally, the boundaries of liver and kidney are placed at the balanced locations of the competition.

The segmentation scheme presented in this paper is a mathematical model for the above process. It is a semi-automatic method because the first step above is achieved by the user providing a label map indicating different targets by different labels. Each subsequent step is handled by an automatic algorithm and is detailed in what follows below.

2.1 Online Feature Learning

Denote the image to be segmented as $I : \Omega \to \mathbb{R}$ where $\Omega \subset \mathbb{R}^d$ is an open set and $d \in \{2, 3\}$. Likewise, the user provided label map is denoted as $L : \Omega \to \mathbb{N} \cup \{0\}$ where 0 indicates background and non-zero positive integers indicate the target object labels. For ease of discussion, in this paper, we assume the distinct labels to be consecutively ranging from 0 to N, an arbitrary positive integer. Moreover, the labeled region can be defined by several "clicks", and does not have to be close to the desired boundary. Next, voxels with the non-zero labels are categorized into different "seed groups" as $G_i = \{\boldsymbol{x} \in \Omega : L(\boldsymbol{x}) = i\}$.

In order to fully utilize the information given by the label map, we note that the seed group not only indicates the location of the target, but also provides some sample voxels contained in it. Hence, instead of making general assumptions on the target characteristics such as brighter/darker than surrounding area, we can learn them in an online fashion. Often times, the image intensity alone is not descriptive enough. Hence, a feature vector is extracted at each voxel, forming a feature image $\boldsymbol{f} : \Omega \to \mathbb{R}^{D_f}$. Subsequently, the segmentation is performed in the feature space. There are many choices for the feature vector such as wavelet coefficients, Fourier descriptors, Hessian matrix, etc. In this paper, we choose local robust statistics [21,2] because they are not sensitive to image noise, and may be computed quickly.

To this end, for each voxel $\boldsymbol{x}$ in the image, we define the feature vector $\boldsymbol{f}(\boldsymbol{x}) \in \mathbb{R}^{D_f}$ by combining several robust statistics derived in a neighborhood $B(\boldsymbol{x}) \subset \Omega$ around $\boldsymbol{x}$. More explicitly, we denote $MED(\boldsymbol{x})$ as the intensity median within $B(\boldsymbol{x})$. In addition, the local intensity range is also an important characteristic, but is sensitive the noise. To address this issue, the distance between the first and third quartiles, namely the inter-quartile range ($IQR(\boldsymbol{x})$), is calculated as the

second feature. Furthermore, the local intensity variance is a good candidate but again it is sensitive to outliers. In contrast, the median absolute deviation (MAD) is much more robust and is computed as $MAD(\boldsymbol{x}) := median_{\boldsymbol{y}\in B(\boldsymbol{x})}(I(\boldsymbol{y}) - MED(\boldsymbol{x}))$. Consequently, we define the feature vector $\boldsymbol{f}(\boldsymbol{x})$ as:

$$\boldsymbol{f}(\boldsymbol{x}) = (MED(\boldsymbol{x}), IQR(\boldsymbol{x}), MAD(\boldsymbol{x}))^T \in \mathbb{R}^3 \tag{1}$$

With the space of feature vectors thus defined, seed groups are now characterized by the probability density function of the feature vectors estimated by:

$$p_i(\boldsymbol{f}) = \frac{1}{|G_i|} \sum_{\boldsymbol{x}\in G_i} K_\eta(\boldsymbol{f} - \boldsymbol{f}(\boldsymbol{x})) \tag{2}$$

where K is the kernel function. In this work, we use the Gaussian kernel. Its variance is chosen to be η times the MAD of the seed group. η is preset to be 0.1, and we have found that this works for all the cases tested.

2.2 Contour Evolution

To simplify the notation, we present the contour evolution in 2D. However it is noted that the method can be easily extended to 3D. In fact, all the experiments in Section 3 are in 3D. First, we denote the family of evolving closed contours as $C_i : [0, 1] \times \mathbb{R}^+ \to \mathbb{R}^2$. Without interactions among contours (interaction is addressed in Section 2.3 below), each contour evolves independently in order to minimize the energy functional:

$$E_i(C_i) := \int\limits_{\boldsymbol{x}\ in\ C_i} (p^c - p_i(\boldsymbol{f}(\boldsymbol{x})))d\boldsymbol{x} + \lambda \int_{C_i} ds \tag{3}$$

where p^c is the cut-off probability density used to prevent the contour leakage [22]. Likewise, $\lambda > 0$ is the smoothness factor. Computing the first variation of E_i and we obtain the flow of C_i:

$$\frac{\partial C_i(q,t)}{\partial t} = [p^c - p_i(\boldsymbol{f}(C_i(q,t))) + \lambda\kappa_i(q,t)]\, \boldsymbol{N}_i(q,t) \tag{4}$$

in which $\boldsymbol{N}_i$ is the inward unit normal vector field on C_i and κ_i is the curvature of the contour.

2.3 Contour Interaction

Although the p^c term in equation (3) helps to prevent contour leakage, in many cases the result is not sufficiently satisfying. Indeed, it often results in the problem that certain regions are over-segmented, while some others are under-segmented. The leakage issue, i.e., making decisions in a transitional region, is sometimes a difficult task even for the human visual system. However, one particular strategy the visual system takes, is to approach the decision boundary from both sides by competition, rather than preventing the leakage from a single

direction. To this end, we enable the interaction amongst the previously individually evolving contours using standard principles from Newtonian mechanics. First, we regard the right hand side of equation (4) as the force applied on the infinitesimal curve segment at the position $C_i(q,t) =: \boldsymbol{p} \in \mathbb{R}^2$. Now with the interaction among curves, such a curve segment will also experience forces from other curves:

$$F_i^{ext}(\boldsymbol{p}) = -\sum_{j \neq i} \int_{C_j} e^{-|\boldsymbol{p}-C_j(w,t)|}(p_j(\boldsymbol{f}(\boldsymbol{p})) - p^c)\boldsymbol{N}_j(\boldsymbol{p})dw \tag{5}$$

Accordingly, the curve flow equation for C_i is now updated as:

$$\frac{\partial C_i(q,t)}{\partial t} = \left[p_i(\boldsymbol{f}(C_i(q,t))) - p^c - \lambda\kappa_i(q,t)\right]\boldsymbol{N}_i(q,t) + F_i^{ext}(C_i(q,t)) \tag{6}$$

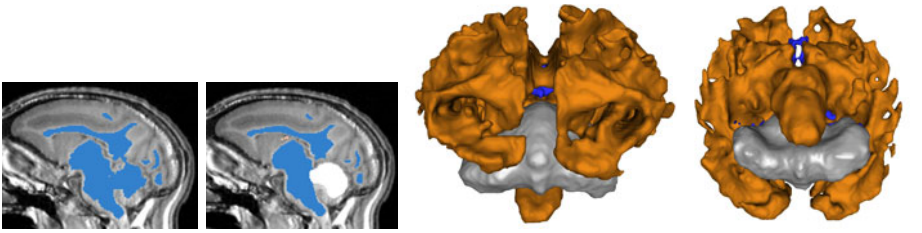
The exponential term controls the "influence range" of the force. When curves are far away, this term reduces the F_i^{ext} effectively to zero. Moreover, using the "sparse field level set" implementation [23], the computation of F_i^{ext} is very efficient. In general, the contour evolution scenario is as follows: At the outset, the contours do not touch each other because the seeds are sparsely scattered in the domain. Thus each F_i^{ext} is approximately 0 and each contour evolves individually. As the evolution proceeds, the contours get closer and the mutual interactions begin to take place. Moreover, they will compete and finally rest at balanced (equilibrium) positions. Throughout the whole process, the contours are governed by the action/reaction principle from mechanics, and will never overlap with each other, which is a necessary feature for multi-object segmentation.

3 Implementation, Experiments and Results

Numerically, the contour evolution is implemented using the sparse field level set method for fast computation and flexibility in contour topology [23]. Moreover, in computing the robust statistics, the neighborhood size $B(\boldsymbol{x})$ is fixed at $3 \times 3 \times 3$. This value was used throughout all of our tests. Similarly, the p^c, λ in equations (6) are respectively fixed at 0.1, 0.3 for all of the tests. In what follows, we demonstrate the application of the proposed method in T1 weighted MR brain imagery and CT abdominal data, to illustrate the algorithm's robustness to the imaging modalities and noise. The results are also quantitatively evaluated.

3.1 Vervet Brain Segmentation

We first test on a T1 weighted MR images of the brain of vervets. In order to highlight the leakage problem as well as how the proposed multi-object scheme solves this problem, initially, only the white matter is segmented. As shown in Figure 1(a), the contour leakage gives a final result that contains not only white matter but also part of cerebellum. However, using the proposed method to segment several related objects gives the result shown in Figure 1(b). It can be seen



(a) (b) (c) (d)

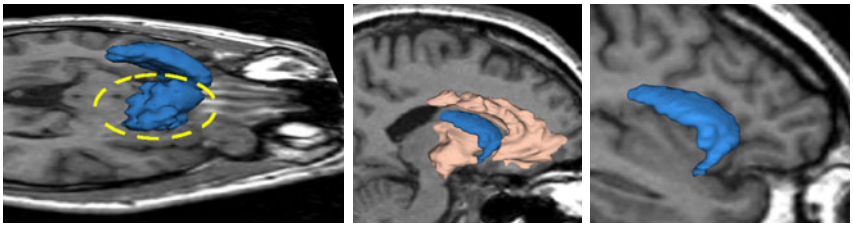
Fig. 1. In Subplot 1(a), we only segment one object (white matter). However, the contour leaks into part of cerebellum and part of brain stem. In 1(b), when segmenting several objects simultaneously, the white label for cerebellum effectively prevents the leakage. 3D plots include posterior 1(c) and inferior 1(d) views. It can be observed that there is no intersection between the surfaces.

that the final labeling of the cerebellum, shown in white, not only fully captures the cerebellum region, but also effectively prevents leakage from intruding into the white matter. Furthermore, we show the 3D views of the multiple segmented objects: white matter, cerebellum, and ventricle. To highlight the region where the contour interaction between the white matter and cerebellum helps prevent leakage, we show the view from both posterior and inferior. It can be observed that there is no intersection between the contours. In particular, the cerebellum contour nicely "pushes" the white matter contour out, and so prevents leakage into the cerebellum.

3.2 Quantitative Analysis for Ventricle and Caudate Nucleus

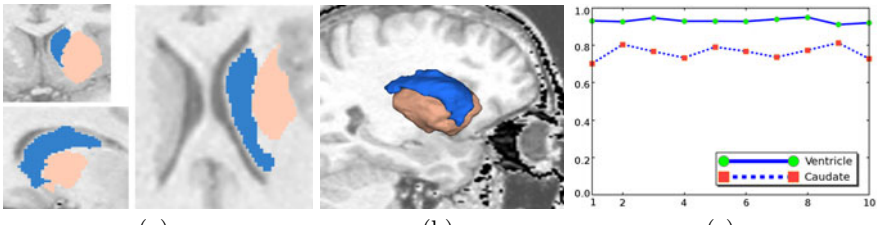
In this second experiment, we extract both the ventricle and the caudate nucleus from MR images and present the results both qualitatively and quantitatively. In the experiment, the caudate nucleus is a difficult object to extract due to the poor contrast with its surrounding tissues. In fact, if we only place seeds in the caudate, we get the result shown in Figure 2(a) where the large leakage is circled. On the other hand, if we also places some seeds around caudate, we also capture some portion of white matter as shown in Figure 2(b) in almond color. Simultaneously, the caudate shape is kept intact and no leakage occurs. The almond part can be discarded because the caudate is the only object of interested and the final result is shown in Figure 2(c).

Performing the same scheme on another subject gives the results in Figure 3(a) and 3(b) where we show both the segmentation and the original image. In addition to the caudate, the method is also applied on ventricle which is an easier segmentation task. In total, we performed 10 tests on different subjects. The Dice coefficients are computed against expert segmentations, and are plotted in Figure 3(c).



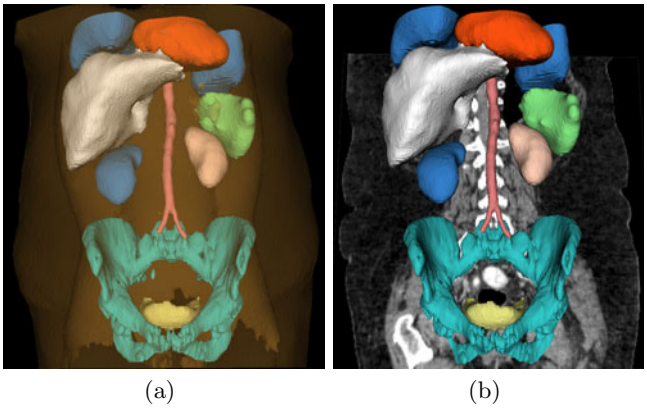
(a) (b) (c)

Fig. 2. If only place seeds in caudate we get segmentation in Subplot 2(a) where the leakage is circled in yellow (viewing from superior-right). After putting some auxiliary seeds in the surrounding tissue we get results in the sagittal view in 2(b) where the caudate shape is kept intact. Discarding the auxiliary region and the caudate is shown alone in 2(c). (Sagittal view from right.)



(a) (b) (c)

Fig. 3. Subplot 3(a) and 3(b) overlay the segmentation results on the original images. The almond region is again auxiliary for preventing leakage. Subplot 3(c) shows the Dice coefficients of segmenting 10 ventricles and caudates, comparing with expert segmentation.



(a) (b)

Fig. 4. Segmentation of heart, two lungs, liver, two kidneys, spleen, abdominal aorta, pelvis, bladder, skin/muscle/fat. The subplot 4(b) removes skin/muscle/fat but overlays the original image.

3.3 Abdominal Organ Segmentation

The proposed algorithm is general purposed and can be used for many different tasks. Indeed, although the previous examples only utilize the multi-object segmentation capability for leakage prevention, in the last experiment, 11 different organs/tissues are extracted from an abdominal CT image. The size of the image is $512 \times 512 \times 204$ and the running time on a machine with 3.0GHz Intel Core 2 Quad CPU and 8G memory is about 8 minutes. The result is shown in Figure 4.

4 Conclusions and Future Work

In this note, we proposed a general-purpose image segmentation scheme for medical data. In particular, the image features are extracted using certain local robust statistics as the segmentation criterion. Subsequently, the object characteristics are learned from the user initialization which is further used to guide the active contour evolution in a variational framework. Furthermore, we incorporate the interactions between the contours into the evolution motivated by simple principles from mechanics. This not only effectively reduces the contour leakage, but also results in a multi-object segmentation scheme without assuming that the union of the segmentation regions is the entire the whole domain.

Future work includes exploring more choices for the image features, such as Fourier/wavelet descriptors. Furthermore, we will incorporate shape priors for the multiple targets. Combined with the contour interaction, this is expected to further improve our results.

References

1. Sonka, M., Hlavac, V., Boyle, R.: Image Processing, Analysis, and Machine Vision, 2nd edn. International Thomson (1999)
2. Pichon, E., Tannenbaum, A., Kikinis, R.: A statistically based flow for image segmentation. Medical Image Analysis 8(3), 267–274 (2004)
3. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. IJCV 1(4), 321–331 (1988)
4. Malladi, R., Sethian, J., Vemuri, B., et al.: Shape modeling with front propagation: A level set approach. IEEE TPAMI 17(2), 158–175 (1995)
5. Kichenassamy, S., Kumar, A., Olver, P., Tannenbaum, A., Yezzi, A.: Gradient flows and geometric active contour models. In: IEEE ICCV, p. 810 (1995)
6. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. IJCV 22(1), 61–79
7. Yezzi, A., Kichenassamy, S., Kumar, A., Olver, P., Tannenbaum, A.: A geometric snake model for segmentation of medical imagery. IEEE TMI 16(2), 199–209 (1997)
8. Yezzi Jr, A., Tsai, A., Willsky, A.: A statistical approach to snakes for bimodal and trimodal imagery. In: IEEE ICCV, vol. 2 (1999)
9. Chan, T., Vese, L.: Active contours without edges. IEEE TIP 10(2), 266–277 (2001)
10. Michailovich, O., Rathi, Y., Tannenbaum, A.: Image segmentation using active contours driven by the bhattacharyya gradient flow. IEEE TIP 16(11), 2787
11. Lankton, S., Tannenbaum, A.: Localizing Region-Based Active Contours. IEEE TIP 17(11), 2029–2039 (2008)

12. Zhu, S., Yuille, A.: Region competition: Unifying snakes, region growing, and bayes/MDL for multiband image segmentation. IEEE TPAMI 18(9), 884 (1996)
13. Brox, T., Weickert, J.: Level set segmentation with multiple regions. IEEE TIP 15(10), 3213 (2006)
14. Vese, L., Chan, T.: A multiphase level set framework for image segmentation using the Mumford and Shah model. IJCV 50(3), 271–293 (2002)
15. Grady, L.: Random walks for image segmentation. IEEE TPAMI 28(11), 17–68
16. Zhao, H., Chan, T., Merriman, B., Osher, S.: A Variational Level Set Approach To Multiphase Motion. Journal of computational physics 127(1), 179–195 (1996)
17. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE TPAMI 22(8), 888–905 (2000)
18. Yan, P., Shen, W., Kassim, A., Shah, M.: Segmentation of neighboring organs in medical image with model competition. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 270–277. Springer, Heidelberg (2005)
19. Yang, J., Staib, L., Duncan, J.: Neighbor-constrained segmentation with level set based 3-D deformable models. IEEE TMI 23(8), 940 (2004)
20. Palmer, S.: Vision science: Photons to phenomenology. MIT Press, Cambridge (1999)
21. Huber, P., Ronchetti, E.: Robust statistics. Wiley-Blackwell (2009)
22. Yang, Y., Tannenbaum, A., Giddens, D., Coulter, W.: Knowledge-based 3D segmentation and reconstruction of coronary arteries using CT images. In: IEEE EMBS, pp. 1664–1666 (2004)
23. Whitaker, R.: A level-set approach to 3D reconstruction from range data. IJCV 29(3), 231 (1998)

Spotlight: Automated Confidence-Based User Guidance for Increasing Efficiency in Interactive 3D Image Segmentation

Andrew Top[1], Ghassan Hamarneh[1], and Rafeef Abugharbieh[2]

[1] Medical Image Analysis Lab, Simon Fraser University
{atop,hamarneh}@cs.sfu.ca
[2] Biomedical Signal and Image Computing Lab, University of British Columbia
rafeef@ece.ubc.ca

Abstract. We present Spotlight, an automated user guidance technique for improving quality and efficiency of interactive segmentation tasks. Spotlight augments interactive segmentation algorithms by automatically highlighting areas in need of attention to the user during the interaction phase. We employ a 3D Livewire algorithm as our base segmentation method where the user quickly provides a minimal initial contour seeding. The quality of the initial segmentation is then evaluated based on three different metrics that probe the contour edge strength, contour stability and object connectivity. The result of this evaluation is fed into a novel algorithm that autonomously suggests regions that require user intervention. Essentially, Spotlight flags potentially problematic image regions in a prioritized fashion based on an optimization process for improving the final 3D segmentation. We present a variety of qualitative and quantitative examples demonstrating Spotlight's intuitive use and proven utility in reducing user input by increasing automation.

1 Introduction

Manual segmentation in 3D medical images is extremely time consuming and tedious. Despite the huge effort often involved, manual segmentation typically suffers from high operator variability and less than ideal results due to user fatigue [11]. Numerous automated algorithms have been developed to aid in image segmentation. Unfortunately, current state-of-the-art fully-automatic algorithms still have difficulty segmenting highly variable shapes such as anatomical structures, hence requiring considerable training resources as well as initialization and fine tuning of unintuitive parameters. Automatic methods cannot guarantee perfect segmentations since the user is not directly involved in the segmentation process. It can thus be rather difficult to correct errors of automatic segmentation, necessitating a final quality assurance pass using some separate interactive segmentation editing tool [8,6].

Semi-automatic methods have been long introduced [11], where the user is factored in to play a larger role in guiding the segmentation process and in correcting errors as they occur. For example, graph cuts [3] and random walker

B. Menze et al. (Eds.): MICCAI 2010 Workshop MCV, LNCS 6533, pp. 204–213, 2011.

algorithms [5,7] can be used for interactive segmentation by specifying object and background seeds. Alternative approaches such as Livewire [2] detect contours in 2D by finding the minimal cost path through user specified seedpoints in the image. Livewire has been shown to be highly reproducible [2], and since the user can see the path forming in real-time, errors can be corrected as they appear. Malmberg et al. [10] attempted to extend Livewire to 3D by removing the restriction of Livewire contours to planes but, unfortunately, the interface used to create the non-planar user-specified contours requires special haptic hardware. Armstrong et al. [1] suggested a 3D analog of a 2D Livewire, namely a "Live Surface", which can be localized in real-time through an optimized graph cuts algorithm. The more recent turtle map 3D Livewire algorithm [12], or T-LW for short, accepts a sparse set of user-specified planar contours as input, which it automatically combines to seed unvisited planes and subsequently generate a new dense set of 3D contours.

A common problem among current interactive segmentation algorithms is that while they generally produce better results as the user provides more input, it is typically not clear to the user what extra input would most improve the segmentation and where additional intervention would be optimal. To address this optimal input problem, we propose an approach that draws the user's attention to low-confidence regions in intermediate segmentations, prodding him/her to provide more information. Effectively, we are automating the process of choosing where additional input should be added, relaxing the burden on the user. In analogy to how theatres use spotlights to focus the audience's attention to an interesting area on the stage, we name our technique 'Spotlight'[1].

2 Method

We build Spotlight on top of an enhanced version of the T-LW algorithm [12]. Although other base methods such as graph cuts and random walker exist and may be adopted (see the Conclusions section), we chose T-LW because it allows for the more natural interaction experience of contouring within 2D slice planes. Users specify seeds along the boundary of the object, enforcing the segmentation to go through it. The reader is referred to the T-LW paper [12] for a detailed explanation of the algorithm. In this work, the T-LW user interaction mechanism remains the same but with the addition of a much needed "suggest plane" feature, which invokes Spotlight to locate a 2D LW contouring plane chosen optimally in regions of maximal segmentation ambiguity, which in turn improves accuracy and speeds up the interactive segmentation process.

In order to enable optimal plane suggestions, which can naturally be in any oblique orientation, we first extend the T-LW algorithm to support user contour specification in arbitrary views as opposed to only the three orthogonal fixed views (section 2.1). We then arm T-LW with our proposed automatic plane suggestion technique which consists of two components; segmentation evaluation (section 2.2) and contour plane suggestion (section 2.3). Three evaluation

[1] A software implementation can be found at http://www.turtleseg.org

criteria based on different metrics create a set of objects called Spotlight Attractors (SAs), which indicate suspected poor segmentation quality regions. Each Spotlight Attractor is a triplet $(\boldsymbol{p}, \boldsymbol{n}, s)$, where $\boldsymbol{p}$ and $s \in [0, 1]$ are the SA's 3D position and strength, respectively, and $\boldsymbol{n}$ is a unit-length direction vector that influences the orientation of the suggested plane. Subsequently, the plane suggestion component determines the 'optimal' high priority plane that passes through the vicinity of the SA positions and conforms to the normals of as many strong SAs as possible.

In summary, a user provides a sparse set of initial contours of arbitrary orientation, and then proceeds to iterate between applying the turtle map algorithm to obtain a 3D Livewire segmentation and examining what Spotlight has flagged as a problematic region for potential intervention. At no point is the user locked to any specific sequence of events, rather the method allows for total freedom in choosing the image planes to be segmented at all times.

2.1 Extending T-LW to Oblique Slices

The T-LW algorithm [12] automatically generates Livewire contours for an unvisited plane $\mathcal{P}_m$ based on the existing sparse set of interactively-seeded 2D contours. A plane $\mathcal{P}_m$ will contain a turtle map, which is a set of line segments resulting from intersecting $\mathcal{P}_m$ with the user-generated contours in other planes not parallel to $\mathcal{P}_m$ (Fig. 1).

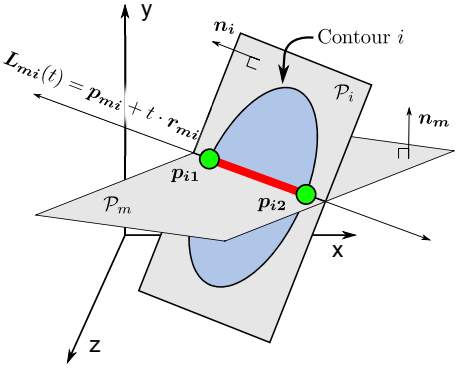


Fig. 1. Computing the turtle map line segments: For contour i (shown encapsulating the blue region in the contour plane $\mathcal{P}_i$), $\boldsymbol{L_{mi}}(t)$, with real parameter t, is the parameterized line formed by intersecting the unvisited plane $\mathcal{P}_m$ with $\mathcal{P}_i$. $\boldsymbol{p_{mi}}$ is a point on $\boldsymbol{L_{mi}}$ and $\boldsymbol{r_{mi}}$ is the direction of $\boldsymbol{L_{mi}}$. The points $\boldsymbol{p_{i1}}$ and $\boldsymbol{p_{i2}}$ are where $\boldsymbol{L_{mi}}$ intersects contour i, and form a (red) turtle map line segment on the turtle map (Fig. 2).

The turtle map encodes the position and ordering of automatically generated Livewire seedpoints in $\mathcal{P}_m$. In order to support oblique contouring slices, we develop a continuous analytical representation of the turtle map instead of the discrete binary bitmap of [12] that is restricted to the pixels in the standard three orthogonal views. We achieve this by taking the line segments produced by intersecting each user contour with $\mathcal{P}_m$ (the red line segment in Fig. 1), and then forming a graph where the vertices are the line segment end points (green) and intersection points (pink), as seen in Fig. 2. The endpoints of the graph represent Livewire seeds on $\mathcal{P}_m$. Note that although $\mathcal{P}_m$ is unvisited by the user, the seeds on $\mathcal{P}_m$ belong to segmentations already confirmed by the user. To allow for fully automatic Livewire

contouring in the unseen $\mathcal{P}_m$, the seedpoints are ordered by choosing an arbitrary initial vertex and then traversing the graph. The next vertex chosen in our graph traversal strategy is the one whose edge forms the smallest oriented angle with the previous vertex (Fig. 2). The automatically generated and ordered seeds on the unvisited plane $\mathcal{P}_m$ are then used to fully automatically livewire the object in $\mathcal{P}_m$. A final 3D segmentation is obtained by choosing a dense set of planes parallel to $\mathcal{P}_m$, and then applying the procedure above to each of them.

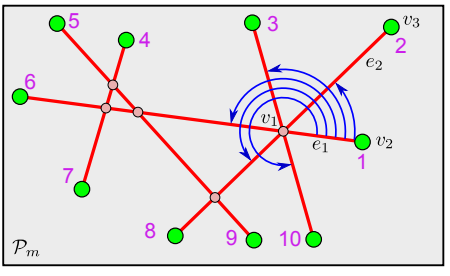


Fig. 2. Line segment end points (Fig. 1) become unordered Livewire seeds. The seeds are ordered as in this example: at intersection vertex v_1 arriving from e_1, the next edge is that which forms the smallest counter clockwise angle with e_1, in this case, e_2. Following e_2 gives the vertex after v_2 as v_3. The purple numbers show the ordering from repeating this procedure.

2.2 Segmentation Assessment

Following the automatic construction of a 3D segmentation, three confidence criteria (i-iii) are evaluated for all automatically-segmented slices.

(i) Contour Edge Strength. Due to potential lack of seedpoints in unvisited slices, the automatically generated contour may incorrectly cut through homogeneous (i.e. edge-free) regions of the object. At such locations, the Livewire local edge cost of the resulting contour (normalized by length) will be high. Hence, for each sample j created by sampling the generated contour at equal arc length intervals, we obtain a position $\boldsymbol{p_j}$ and a corresponding normal $\boldsymbol{n_j}$ (normal to the contour and lying in the contour's plane), which are then used to generate a SA: $(\boldsymbol{p_j}, \boldsymbol{n_j}, f_{LWj})$. Here, f_{LWj} is the Livewire local edge cost at point j, based on a combined Canny edge, Laplacian of Gaussian, gradient magnitude and gradient direction terms [12].

(ii) Contour Stability. The Livewire algorithm calculates the globally minimal path between two points, but gives no evidence of other near-optimal paths. Inspired by the k-shortest paths problem [13], we detect such paths by perturbing single edges in the shortest path (Fig. 3) to produce a set $\mathcal{D}$ of perturbed paths. Our contour instability value is calculated, at each point k on the contour, as the maximum Euclidean distance between the position $\boldsymbol{p_k}$ of k and each perturbed path, given by $d_k = \max_{R\in\mathcal{D}}(\min_{\boldsymbol{p}\in R}(|\boldsymbol{p_k} - \boldsymbol{p}|_2))$. We then create the SA as $(\boldsymbol{p_k}, \boldsymbol{n_k}, u(d_k))$, where $\boldsymbol{p_k}$ and $\boldsymbol{n_k}$ are as before (in (i)) and $u(x)$ is a monotonically increasing function that assigns a strength ($\in [0, 1]$) to the distance. In our implementation, $u(x)$ normalizes the distance with respect to the largest distance between points in the existing contours.

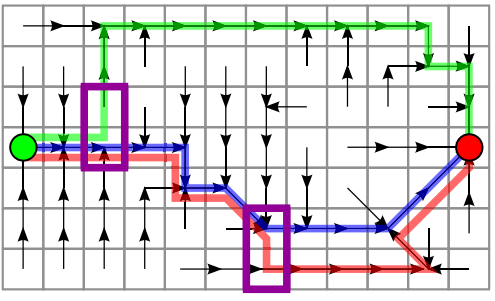


Fig. 3. Path perturbation example. New paths (green and red) formed through perturbations of the shortest path (blue). The path perturbation edges are outlined in purple. The thin black arrows represent the shortest path tree from the source (green dot) to the destination (red dot). Note how a local change can result in a substantially different path (green).

(iii) Turtle Map Connectivity. This metric detects unvisited planes that lack the seeds necessary to reveal the cross section of the object. A turtle map containing non-intersecting line segments would provide only 2 seedpoints, typically resulting in a poor segmentation. For each non-intersecting line segment h in an unvisited plane with normal $\boldsymbol{n_h}$ within that plane, we sample points i along h at a resolution proportional to that of the source image. Finally, for each i, we generate a SA as $(\boldsymbol{p_i}, \boldsymbol{n_h}, 1)$.

2.3 Spotlight: Automated Slice Plane Suggestion

In order to produce an optimal slice plane on which the user should focus his/her attention, we devise an objective function that assigns a cost to a plane given the set of generated SAs. At a high level, we derive a cost for the plane with respect to each SA, and sum these costs over all SAs. Given a plane $\mathcal{P}$ with normal $\boldsymbol{n_\mathcal{P}}$ and offset from origin $d_\mathcal{P}$, the cost of a suggested plane $\mathcal{P}$ is given by

$$E(\mathcal{P}) = \frac{\sum_{i \in SA} s_i E_{SA}(i, \mathcal{P})}{\sum_{i \in SA} s_i} \tag{1}$$

where s_i is the strength of attractor i. $E_{SA}(i, \mathcal{P})$ is the cost contribution associated with the ith SA, which is low if the plane $\mathcal{P}$ is near and parallel to the SA's normal. Therefore, E_{SA} is defined as

$$E_{SA}(i, \mathcal{P}) = q_\mathcal{P}(i) + (1 - q_\mathcal{P}(i)) L(d(\boldsymbol{p_i}, \mathcal{P})). \tag{2}$$

Here, $q_\mathcal{P}(i) = (\boldsymbol{n_i} \cdot \boldsymbol{n_\mathcal{P}})^2$ describes how similar the plane's normal is to the Spotlight Attractor's, where the more perpendicular the normals are, the lower the value of $q_\mathcal{P}(i)$ and hence a value of E_{SA} closer to $L(d(\boldsymbol{p_i}, \mathcal{P}))$. We set $d(\boldsymbol{p}, \mathcal{P}) = |\boldsymbol{p} \cdot \boldsymbol{n_\mathcal{P}} - d_\mathcal{P}|$. $L : [0, \infty) \rightarrow [0, 1)$ is a logistic shaped function used to reduce the influence of points far from $\mathcal{P}$ on the optimizer.

We use gradient descent to minimize E with respect to $\mathcal{P}$. We iterate over multiple initialization planes, each passing through at least one SA, and pick the plane with the lowest minimum.

3 Results

3.1 Implementation Details

We have chosen the algorithm parameters empirically and kept them fixed during all the following experiments, showing that these fixed parameters are robust to a wide variety of image modalities and data. The main parameters used in the Spotlight-augmented segmentation process are those inherited from the T-LW algorithm. Spotlight-only parameters include the gradient descent parameters, as well as the relative weightings between the evaluation metrics. We keep the evaluation metric relative weightings equal. For the gradient descent parameters, we have found a good quality-performance balance at 40 re-initializions, with 80 steps performed each time. Increasing the gradient descent steps beyond these values increases the computation time only; the increase in suggestion plane quality is negligible.

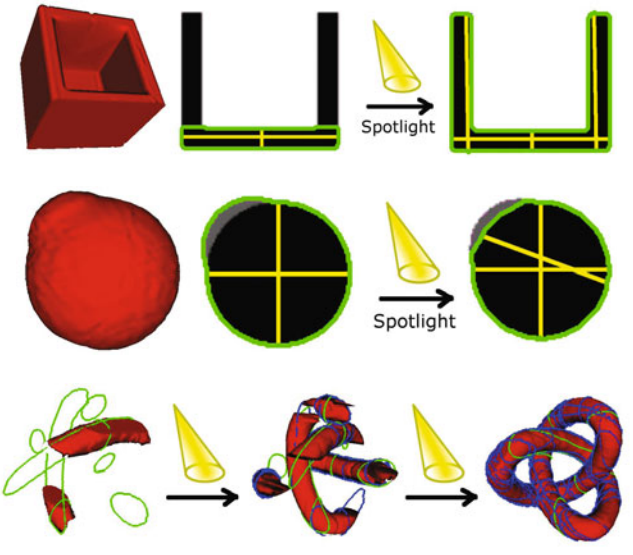


Fig. 4. Spotlight discovering segmentation mistakes. Note how in the top row the sides (vertical black bars) of the box is segmented properly, how the bulge is correctly excluded from the sphere in the second row, and Spotlight's proper handling of complicated non-spherical topology in the third row (see text for details).

3.2 Synthetic Tests

Fig. 4 (row 1) illustrates an open box test example (left) where the initial turtle map cuts through the box walls (middle), creating high Livewire costs that are detected and fixed by Spotlight (right). Fig. 4 (row 2) shows a sphere with an unwanted bump caused by a strong nearby edge (left), creating high path

stability costs (middle) causing Spotlight to suggest a slice plane that removes the ambiguity about where the Livewire path should follow (right). Fig. 4 (row 3) shows Spotlight guiding the segmentation of a knot, where starting with only 2 user-chosen slice planes (green) on the left, we follow numerous suggestions from Spotlight (blue) to fully segment the knot, seen on the right.

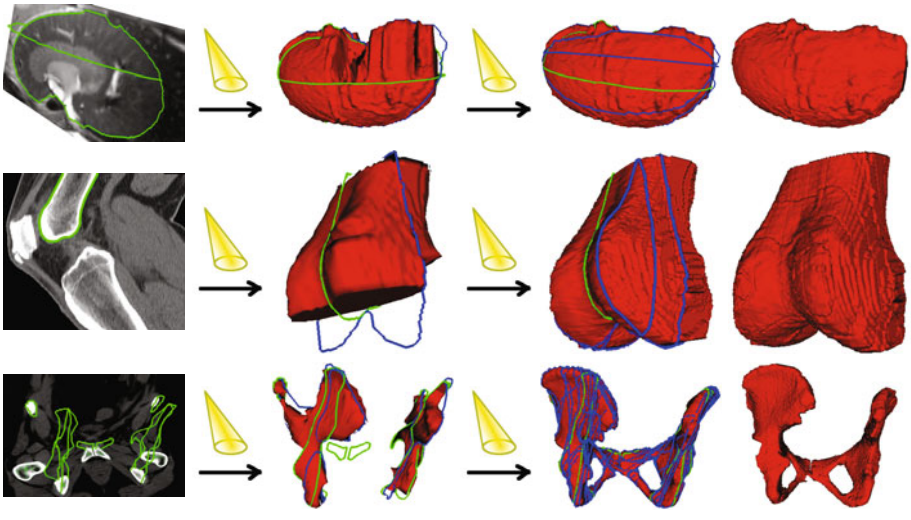


Fig. 5. Segmentation results on real data. (Column 1) Initial contours (green) overlaid on image data slices of (top to bottom) a mouse kidney (MR), femur (CT) and pelvis (CT), respectively. Contours were provided on 2, 1 and 4 different initial planes, respectively. (Column 2) Intermediate result showing some added Spotlight contours (blue) and surface rendering mid-way through the Spotlight suggestion process; after 1, 1, and 3 iterations, respectively. (Column 3) Final surface rendering with contours; 4, 4 and 30 iterations of Spotlight were required, respectively. (Column 4) Surface rendering of final segmentation.

3.3 Real Medical Data Tests

Fig. 5 shows our tests on real medical images. In all examples, an initial set of contours (green) is provided by the user, who proceeds to contour Spotlight suggested planes. We emphasize Spotlight's contribution to these segmentations by coloring blue all contours whose plane was chosen automatically. Fig. 5 (row 1) shows Spotlight aiding the user in segmenting a mouse kidney in a magnetic resonance (MR) image. The data is noisy and contains weak edges, resulting in low Livewire local path costs and this is detected by Spotlight. Fig. 5 (row 2) shows a femur being segmented in a computed tomography (CT) image. Note that Spotlight selects planes that best outline object features such as the lateral and medial condyle. The CT pelvis results in Fig. 5 (row 3) demonstrate our algorithm's ability to navigate real anatomy with non-spherical topology. While Spotlight at first concentrated on correcting errors in the iliam, eventually

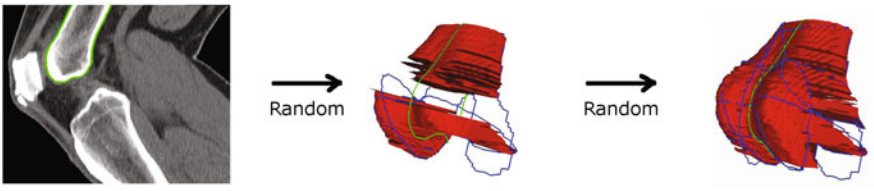


Fig. 6. Segmentation guided by random plane suggestions. Both segmentations were initialized with the same single contour as used in Fig. 5 row 2. The middle image shows the segmentation after 4 randomly chosen contour planes. The right image is after 10 randomly chosen planes. Here, the blue contours are randomly chosen.

segmentation confidence increased in that region and Spotlight began focusing on the smaller ischium region.

As a baseline verification of Spotlight, we have tested it against random plane selection. Fig. 6 shows the progress of a knee femur segmentation (from the same image used in Fig. 5, row 2) guided by random plane suggestions. As expected, Fig. 6 shows significantly poorer random plane selection results compared to Spotlight (Fig. 5, row 2).

3.4 User Study Tests

We have performed a user study in order to quantify the efficiency increase provided by Spotlight over the basic T-LW. In the study, 8 users were asked to segment a synthetic open box (Fig. 4 (row 1)) and a liver in a CT image with provided ground truth from the SLIVER07 MICCAI Grand Challenge [4]. For each image, 4 users were instructed to perform the segmentation without using Spotlight, while the other 4 were instructed to use Spotlight. Users were asked to stop segmenting when they deemed the resulting segmentation accurate. The accuracy of the intermediate 3D segmentation was subsequently evaluated offline after each added contour. The quantitative results of the study are shown in Fig. 7. Fig. 7 shows the increase in Average Dice Similarity coefficient (DSC) vs. number of mouse clicks, number of contours segmented, and time spent. Without Spotlight, the subjects had difficulty determining where to contour next, often choosing to contour planes that did not improve the segmentation at all. The box image was interesting in that there exists a minimum number of contours required for the T-LW algorithm to segment it completely, yet users were typically unable to intuitively determine this best solution. Compared to the original T-LW algorithm in the box image, Spotlight required only 30% of the time and 64% of the contours. For the liver image, total segmentation time was reduced by 35%. Notice also that Spotlight is more consistent in improving the segmentation, exhibiting much smaller standard deviation across users in the DSC.

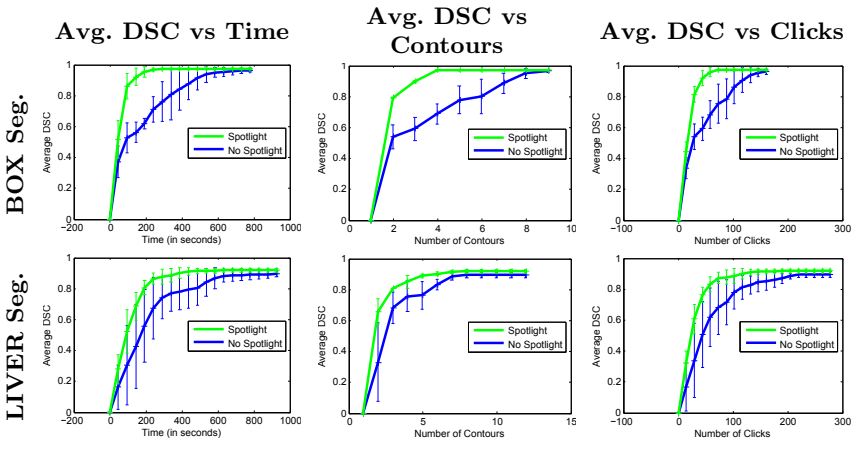


Fig. 7. User study results. Each graph compares Spotlight (green) to the original T-LW (No Spotlight, in blue). In the top row, users were asked to segment the synthetic open box shape of Fig. 4 (row 1). In the bottom row, users were asked to segment the liver in a CT image. The first, second and third columns show how the segmentation accuracy improves as a function of time spent, number of contours added, and number of mouse clicks, respectively. The error bars represent the standard deviation among the 4 user results.

4 Conclusions

We presented Spotlight, a novel confidence-based user guidance technique for increasing efficiency of interactive 3D segmentation. Spotlight's intuitive mechanism automatically draws the user's attention to potential problematic regions by highlighting image planes where user input is very likely to improve the segmentation. To quantify confidence in a segmented contour, we minimized an objective function comprising three metrics reflecting contour edge strength, contour stability and object connectivity. We demonstrated Spotlight's successful application for the segmentation of illustrative synthetic volumes and complex real 3D medical images.

Our current implementation was based on the T-LW algorithm [12], which we extended with an analytical formulation for the turtle map. This enabled the necessary support for using arbitrarily directed oblique planes that may be suggested by Spotlight. The Spotlight strategy, however, is not restricted to this Livewire framework and can be incorporated in to other interactive segmentation algorithms, such as random walker with precomputation [7] or graph cuts [9]. We leave these extensions to future work. Our algorithm can also easily support additional confidence measures to augment the current SAs. Part of our future work will focus on incorporating region-based reliability metrics. Furthermore, we are currently investigating the formulation of Spotlight under an active learning

framework. Finally, we will focus on further testing of Spotlight in various applications and performing extensive validation experiments on real data through additional user studies.

References

1. Armstrong, C.J., Price, B.L., Barrett, W.A.: Interactive segmentation of image volumes with live surface. Computers and Graphics 31(2), 212–229 (2007)
2. Barrett, W.A., Mortensen, E.N.: Interactive live-wire boundary extraction. Medical Image Analysis 1, 331–341 (1997)
3. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient N-D image segmentation. International Journal of Computer Vision 70(2), 109–131 (2006)
4. Heimann, T., et al.: Comparison and evaluation of methods for liver segmentation from CT datasets. IEEE Transactions on Medical Imaging 28(8), 1251–1265 (2009)
5. Grady, L.: Random walks for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(11), 1768–1783 (2006)
6. Grady, L., Funka-Lea, G.: An energy minimization approach to the data driven editing of presegmented images/Volumes. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 888–895. Springer, Heidelberg (2006)
7. Grady, L., Sinop, A.K.: Fast approximate random walker segmentation using eigenvector precomputation. In: IEEE Computer Vision and Pattern Recognition, pp. 1–8 (2008)
8. Kang, Y., Engelke, K., Kalender, W.: Interactive 3D editing tools for image segmentation. Medical Image Analysis 8, 35–46 (2004)
9. Kohli, P., Torr, P.H.S.: Measuring uncertainty in graph cut solutions - efficiently computing min-marginal energies using dynamic graph cuts. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 30–43. Springer, Heidelberg (2006)
10. Malmberg, F., Vidholm, E., Nyström, I.: A 3D live-wire segmentation method for volume images using haptic interaction. In: Kuba, A., Nyúl, L.G., Palágyi, K. (eds.) DGCI 2006. LNCS, vol. 4245, pp. 663–673. Springer, Heidelberg (2006)
11. Olabarriaga, S., Smeulders, A.: Interaction in the segmentation of medical images: a survey. Medical Image Analysis 5, 127–142 (2001)
12. Poon, M., Hamarneh, G., Abugharbieh, R.: Efficient interactive 3D livewire segmentation of complex objects with arbitrary topology. Computerized Medical Imaging and Graphics 32, 639–650 (2008)
13. Yen, J.Y.: Finding the K shortest loopless paths in a network. Management Science 17, 712–716 (1971)

Automated Segmentation of 3D CT Images Based on Statistical Atlas and Graph Cuts

Akinobu Shimizu[1], Keita Nakagomi[1], Takuya Narihira[1], Hidefumi Kobatake[1], Shigeru Nawano[2], Kenji Shinozaki[3], Koich Ishizu[4], and Kaori Togashi[4]

[1] Tokyo University of Agriculture and Technology, Tokyo, Japan
simiz@cc.tuat.ac.jp
[2] International University of Health and Welfare, Tokyo, Japan
[3] National Kyusyu Cancer Center, Fukuoka, Japan
[4] Kyoto University, Kyoto, Japan

Abstract. This paper presents an effective combination of a statistical atlas-based approach and a graph cuts algorithm for fully automated robust and accurate segmentation. Major contribution of this paper is proposal of two new submodular energies for graph cuts. One is shape constrained energy derived from a statistical atlas based segmentation and the other is for constraint from a neighbouring structure. The effectiveness of the proposed energies was demonstrated using a synthesis image with different errors in shape estimation and clinical CT volumes of liver and lung.

Keywords: segmentation, CT, shape prior, neighbour constraint, graph cuts, statistical atlas.

1 Introduction

Statistical atlas-based segmentation of 3D medical images has intensively studied because it provides a framework of effective utilization of shape priors inherent in the target anatomical structures. Especially, statistical shape models (SSMs) [1] have been successfully utilized for automated segmentation [2]-[5]. One problem of this approach is that segmentation accuracy becomes insufficient for shapes that are not covered by training datasets. In order to overcome this problem, subsequent free-form surface fitting [6] and hierarchical refinement using multi-level SSMs [7] were proposed. However, these methods often become unstable partly because of the local minima problem in non-linear optimization involved in these methods.

One of recent noteworthy developments in image segmentation is graph cuts approach [8], in which the global minimum for certain types of energy functions is guaranteed. Previous efforts on incorporating shape priors into the graph cuts approach are mainly classified into the two categories; general constraints such as ellipse and star-shape [9]-[11] and specific constraints such as a user defined scaled rigid template [12]. The shape priors provided by statistical atlas are categorized between the above-mentioned general and specific constraints. Effective combination of the two strong approaches, that is, statistical atlas-based and graph cuts approaches, is desirable for robust and accurate segmentation.

B. Menze et al. (Eds.): MICCAI 2010 Workshop MCV, LNCS 6533, pp. 214–223, 2011.

One simple approach to the combination is using probabilistic atlas and distance histogram of the target organ as shape priors [13],[14]. However, rather than using probabilistic atlas, we start from segmentation result by existing atlas-based segmentation methods, which have already been demonstrated in not a few papers [1]-[7] that approximated organ regions are successfully obtained in most cases. Therefore, the remaining important problem is to refine the approximated organ regions so as to obtain final complete results. Unlike the previous methods for the refinement which suffer from the local minimum problem [6],[7], we take the graph cuts approach, which is expected to provide accurate refinement in a stable manner because of the guarantee of global optimization. In this paper, we formulate an automated procedure of a subsequent refinement process of approximated organ regions obtained by a statistical atlas-based method. Main contributions of our work are as follows.

- Derivation of a newly defined submodular energy function based on constraints from surface normals of an approximated shape and a neighbouring structure so as to become robust against deviation of an approximated shape from its true shape, and pathological changes in an organ.
- Achievement of effective combination of a statistical atlas-based approach and a graph cuts algorithm for fully automated robust and accurate segmentation.
- Validations using a synthetic image with different errors in shape estimation and clinical 3D CT images for segmentation of the liver and lung.

In the following, the details of methods and results are described.

2 Proposed Segmentation Framework

2.1 Graph Cuts

Graph cuts [8] formulates a segmentation problem as an energy minimisation problem. The goal is to find a set of labels $A = (A_1, A_2, \ldots, A_p \ldots, A_{|P|})$ that minimises an energy E(A) given by

$$E(A) = \lambda R(A)+B(A) = \lambda \Sigma_{p\in P} R_p (A_p) + \Sigma_{\{p,q\}\in Np} B_{p,q}\, \delta_{Ap\neq Aq}. \quad (1)$$

where the set P is a set of voxels in a 3D volume, and the energy $R_p(A_p)$ is a matching cost, or t-link cost, of a graph, assigning label $A_p \in L$ to p. The symbol A_p is an element of label set $L = \{1,0\}$; 1 is object and 0 is background. This cost is defined by a negative log likelihood of CT values, where a probability density function of each class is assumed to be a normal distribution with parameters estimated by an EM algorithm. The set Np is a set of voxels in the 6-neighbourhood of p, and the function δ is 1 if $Ap \neq Aq$ and 0 otherwise. The energy $B_{p,q}$ is a n-link cost of labeling the pair p and q with labels $Ap \neq Aq \in L$. Here, the coefficient λ is a constant value balancing the two costs. Detailed explanation of each energy term can be found in [8].

2.2 Proposed Energies for Graph Cuts

This study extends the above energy function to incorporate constraints from a shape prior and a neighbour structure as follows.

$$E(A) = \lambda \Sigma_{p \in P} \{ R_p(A_p) + NB_p(A_p) \} + \Sigma_{\{p,q\} \in Np} \{B_{p,q} + S_{p,q}\} \delta_{Ap \neq Aq}. \quad (2)$$

The energy $S_{p,q}$ is a proposed shape-based energy for n-link of a graph. Once shape of an object is estimated by a statistical atlas-based process (see 2.3), the energy is computed using a signed distance $\phi(p)$ from boundary of an estimated shape. Here, in a signed distance function, an object voxel has negative distance and a background voxel has positive distance. The energy $S_{p,q}$ is defined by the following equation where θ is an angle between a gradient vector of $\phi(p)$ and a vector connecting points p and q.

$$S_{p,q} = \mathrm{sqrt}\{[1 - \cos(\theta)]/2\}. \quad (3)$$

This energy encourages an n-link between p and q to be cut when the direction of a vector connecting p and q is similar to that of a gradient vector of $\phi(p)$ (see Figure 1). Consequently surface normals of an extracted region tend to be parallel to those of an estimated shape, resulting in high similarity between the extracted region and the estimated shape. A notable feature of this energy is that it is insensitive against deviation of an estimated shape from its true shape. Let us suppose that in Figure 1 the boundary C is a boundary of an estimated shape and the boundary C' is that of a true shape. Since gradients of signed distances of the boundary C are nearly identical to those of the boundary C', C' can be extracted when minimizing the shape energy. In section 3.2, we will further discuss robustness against deviation using a synthetic image.

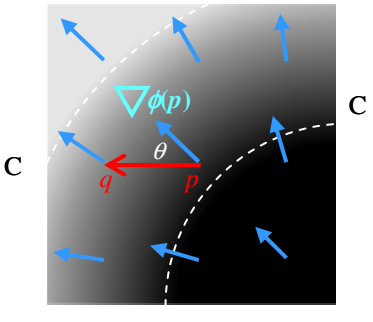


Fig. 1. Gradient vectors of $\phi(p)$ and an angle θ between a gradient vector of ϕ (p) and a vector connecting points p and q. The boundary C is a boundary of an estimated shape and the boundary C' is a true boundary that is similar to C in shape but differs in location.

The energy $NB_p(A_p)$ is a neighbour constrained energy to correct undesirable change of $R_p(A_p)$ mainly caused by pathology. A good example of neighbour structures is a rib when segmenting a lung with diseases. Unfavorable changes by pathology can be suppressed by a prior knowledge that lung fields are surrounded by ribs. The concept of this energy is general and can be applicable to many segmentation problems. However, concrete definition depends on a problem to be solved. In this study we focused on segmentation of a lung field with large pleural effusion. A neighbor constrained energy is defined by distance D_{rib} from dorsal ribs as follows.

$$\begin{aligned} NB_p\,(A_p) &= 0 \quad (if\ A_p = 1\ \text{'obj'}) \\ NB_p\,(A_p) &= D_{rib} \quad (if\ A_p = 0\ \text{'bkg'}). \end{aligned} \tag{4}$$

The dorsal rib can be extracted by a simple binarization process followed by a morphological operation, or a closing operation to fill gaps and holes. The distance D_{rib} is computed in the region surrounded by the extracted dorsal ribs (D_{rib} = zero for outside of the region). The energy of equation (4) is expected to increase the cost of t-link when A_p = 'bkg', which encourages pleural effusion to be classified as lung.

2.3 Statistical Atlas Based Shape Estimation

Shape estimation is important for ensuring that an extracted organ's shape is natural and consistent with individual anatomy. Leventon et al. proposed a level set distribution model (LSDM) based shape estimation in their segmentation process to drive a level set function toward the shape [3]. They maximised a posteriori probability given a level set function and gradient of an input image to estimate the patient specific shape. The objective function was designed for an iterative evolution process of a level set function. Our shape estimation process was inspired by their work but differs in an objective function. We employed a statistical atlas-based shape estimation process that performs maximum a posteriori probability (MAP) segmentation [15] followed by a LSDM based shape fitting. The fitting process finds the most similar shape in an eigen shape space with an extracted region by MAP segmentation. Actually it maximizes a Jaccard Index (J.I.) between an extracted 3D region and shape in a discretized eigen space of LSDM in an exhaustive manner. The most similar 3D shape model is forwarded to the graph cuts algorithm and used as a shape prior.

3 Experimental Evaluations and Comparisons

3.1 Materials

Synthetic Image: A 2D synthetic image of Figure 2(c) was used for performance validation and comparison with Freedman's energy [12]. A four neighborhood system was employed for this experiment only. The image includes not only an object (Figure 2(a)) but also noises consisting of six structural noises (Figure 2(b)). The structural noises mimic lesions and vessels with different contrast in an organ, as well as an additive Gaussian noise ($N(0,20^2)$). Radius of the object in Figure 2(a) is 40 [pixel] on average and amplitude A is 5 [pixel], respectively. Distance D is displacement of an object in a shape template that was used as an estimated shape. In this experiment, we changed amplitude A and distance D of the shape template to realize various deviation from a true shape.

Clinical CT volumes: We performed two segmentation experiments, or liver and lung segmentation from non-contrast CT volumes. It is worthy to note that all processes, or MAP, LSDM based shape estimation and graph cuts, are 3D automated processes. A 6-neighborhood system was employed for the graph cuts. First experiment was liver segmentation using 20 cases with liver metastases. Second experiment

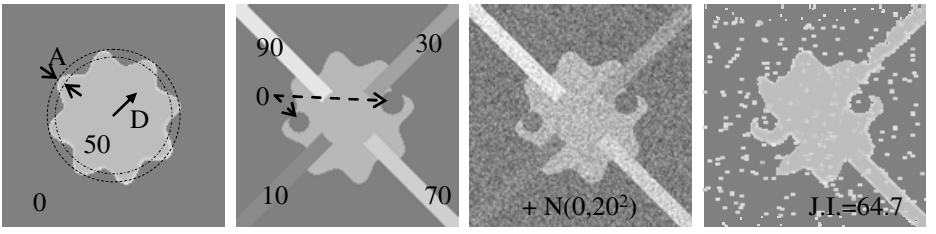


Fig. 2. Illustrations for a synthetic image (150x150). (a) an object label, (b) with structural noises, (c) with Gaussian noise, and (d) a result of graph cuts without a shape prior. Numerals in the figures (a) and (b) are gray values of objects.

was lung segmentation using 97 cases with pulmonary diseases, such as pleural effusion. The performance of the algorithms was assessed by a cross validation test, where training cases for designing a LSDM and deciding parameters of the algorithms, such as λ and σ of $B_{p,q}$ [8], were separated from testing cases.

3.2 Results

Synthetic Image: Primary aim of this experiment is to test robustness of the proposed shape energy $S_{p,q}$ against the deviation of an estimated shape from the true one. We prepared a graph cuts algorithm with the energies of equation (1) plus $S_{p,q}$ and estimated shapes, or shape templates with various deviations.

Figure 3 shows Jaccard Index (J.I.) between extracted regions and a true one, when changing the two parameters, or amplitude A and displacement D of an object in a shape template, from 0 to 15, respectively. Figure 4 presents several segmentation results, each of which corresponds to an arrow with an alphabet in Figure 3.

These figures suggested that the proposed segmentation was accurate and insensitive to the deviation of the estimated shape from the true one. In contrast the performance of Freedman's energy was seriously decreased as the deviation became large. Since they used an unsigned distance function $|\phi(p)|$, the boundary of the segmented object lied near the boundary of the estimated shape, which was typically observed in the Figure 4(e) and (f). Average J.I. over all 256 combinations (= 16x16) was 94.3% for the proposed energy and 89.1% for the Freedman's energy. The difference was statistically significant ($p<0.01$. Wilcoxon) .

Clinical CT volumes: First experiment was liver segmentation which employed $S_{p,q}$ as a shape energy. Figure 5 (a) and (b) are slices of a non-contrast CT volume and extracted regions by a MAP based segmentation. The MAP based segmentation performed relatively well. However, the result contained false positives of surrounding tissues as indicated by an arrow due to weak contrast of boundary and high similarity in CT value. Such error made the shape estimation task difficult. The boundary estimated by a LSDM trained from 10 CT volumes was drawn in white in Figure 5(b). As you can see, the approximation was rough due to the error in the MAP based segmentation. However the proposed subsequent graph cuts algorithm succeeded in the segmentation (see Figure 5(c)), while not for Freedman's energy. Figure 5(d) contained over-extracted surrounding tissues that lied near the estimated shape boundary in Figure 5(b).

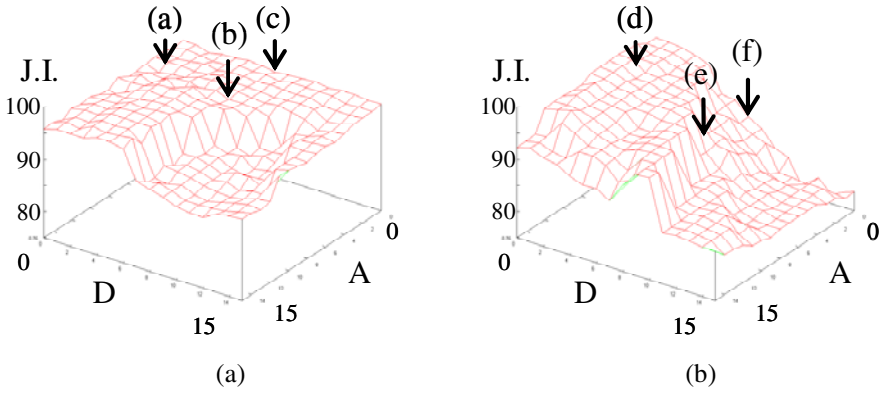


Fig. 3. Segmentation performance for Figure 2(c). (a) proposed shape energy and (b) Freedman's shape energy. Each arrow corresponds to a result of Figure 4.

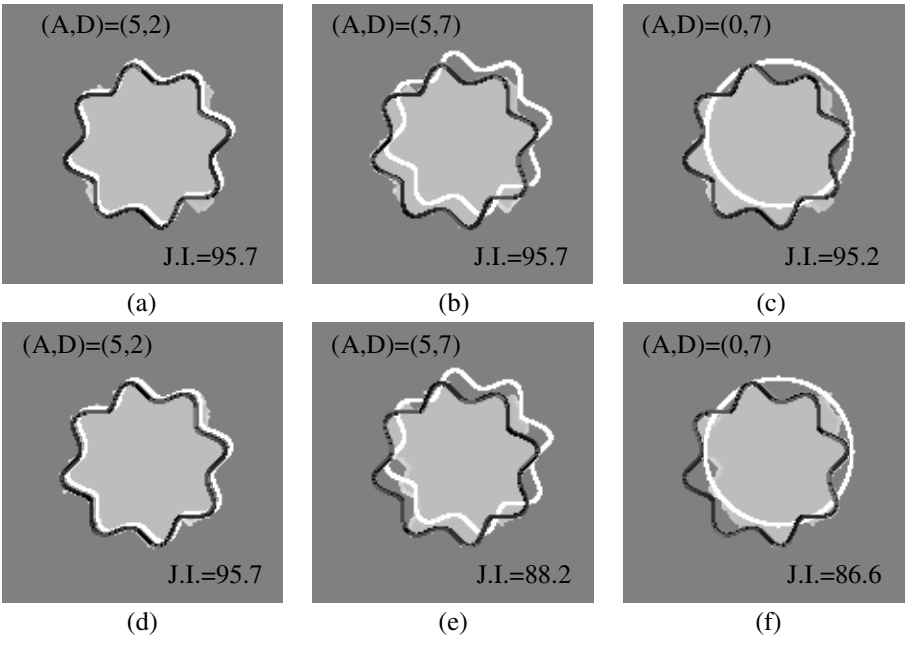


Fig. 4. Segmentation results of Figure 2(c). (a)-(c) proposed shape energy and (d)-(f) Freedman's shape energy. Black lines show true boundaries and white lines are boundaries of shape templates.

The performance was evaluated by a cross validation test using 20 CT volumes. The dataset was divided into two equal-sized subsets, one of which was used to train a LSDM and decide parameters of graph cuts algorithms, and the other was for validation test. In the second round, the roles of the two subsets were exchanged and

performance was averaged over the two rounds. True boundaries were defined by one of the authors and approved by a radiologist. Average J.I. over all data of the MAP based segmentation, graph cuts algorithms without a shape energy and with Freedman's energy were 78.9%, 89.8% and 91.8%, respectively. The highest performance was achieved by the proposed energy whose average J.I. was 93.2%. Wilcoxon test told us that differences between any combination of two algorithms were statistically significant ($p<0.01$).

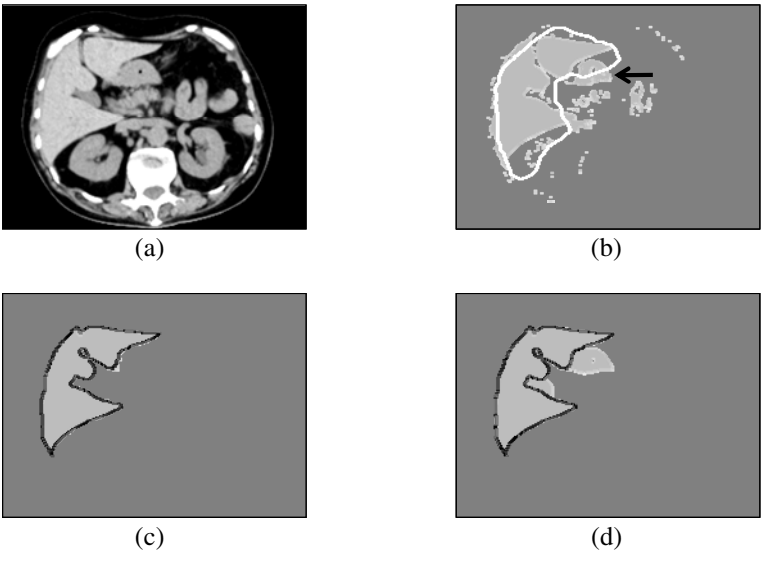


Fig. 5. (a) a non-contrast CT image, (b) MAP based segmentation with an estimated shape (a white line), (c) proposed segmentation result, and (d) segmented region based on the Freedman's energy. Black lines of (c) and (d) are true boundaries.

Second experiment was lung segmentation. The proposed algorithm used both the energies $S_{p,q}$ and NB_p (A_p) because of large pleural effusion. Figure 6 presents an original image and segmentation results. A graph cuts algorithm without a shape energy misclassified both the pleural effusion and hilum vessels (see Figure 6(b)). The shape energy computed from an estimated shape (a white line in Figure 6(c)) improved the mis-segmentation of hilum vessels as shown in Figure 6(c). However the error in pleural effusion was remained unsolved. The proposed neighbour constrained energy corrected the t-link cost and the pleural effusion was classified into the lung field correctly (Figure 6(d)).

A cross validation test using 97 CT volumes was performed to validate the three graph cuts algorithms, or an algorithm without a shape prior, that with the shape energy and that with both the shape energy and the neighbour constrained energy. In that experiment, we divided 97 CT volumes into 48 and 49 CT volumes, and carried out two rounds experiment as in the liver segmentation.

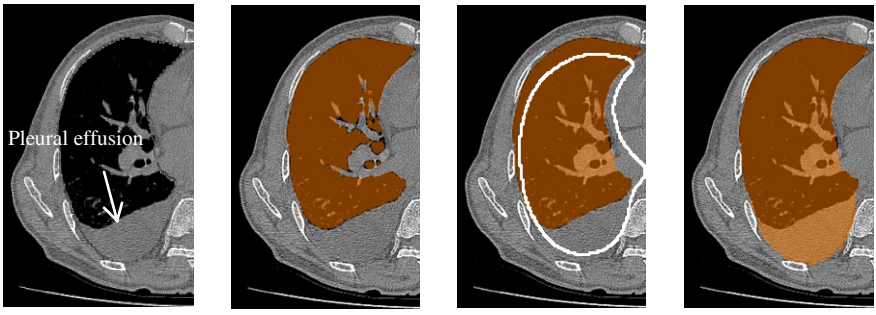


Fig. 6. (a) an original image with a true boundary, (b) graph cuts segmentation without a shape energy, (c) graph cuts segmentation with $S_{p,q}$ and (d) proposed graph cuts segmentation with $S_{p,q}$ and NB_p (A_p)

The performance was evaluated by not only J.I. but also distance between an extracted surface and a manually delineated surface. Average J.I. of graph cuts without a shape energy, with the proposed shape energy and with both the shape energy and the neighbour constrained energy were 95.7%, 97.4% and 97.7%, respectively. We performed Wilcoxon test to confirm the statistical difference in performance between any two algorithms out of three and found that the differences were significant with the risk $p < 0.01$. Although the improvement by the neighbour constrained energy is small, it increased the J.I. for 75 cases out of 97 cases, resulting in statistical difference. The average distance over all testing data was decreased by the graph cuts with the shape energy from 2.30 [voxel] to 0.627 [voxel]. Here the voxel size is 0.683 [mm] on average. The addition of the neighbour constrained energy decreased the error to 0.464 [voxel]. The differences between the three distance distributions were statistically significant (Wilcoxon test, $p < 0.01$).

4 Discussion

This paper proposed a new segmentation algorithm that combined a statistical atlas-based shape estimation with a graph cuts algorithm to incorporate constraints from the estimated shape and a neighbour structure. A salient feature of the proposed shape constrained energy is robustness against the deviation of an estimated shape from the true one, which was demonstrated by the experiments using a synthetic image as well as liver segmentation using 20 CT volumes. Additionally we would like to enhance the robustness against the well-known problem of a LSDM [4][5]. Since the space of signed distance functions is not a linear space, an estimated surface by LSDM becomes smoother than true one, leading to large deviation. Our proposed energy could compensate the disadvantage of the LSDM.

Constraint from a neighbour structure derived a new energy that focused on surrounding reliable structure. In this paper, we proposed a rib constrained energy for lung segmentation with large pleural effusion. The cross validation test using 97 CT volumes told us that the performance was significantly improved by the proposed

neighbour constrained energy. Computational time of a max-flow algorithm to optimize the energy function was roughly 1 to 2 minutes (Xeon(TM) 3.6 GHz x 2).

Finally, several limitations of the proposed segmentation algorithm and future works warrant attention. First, as is found in the Figure 3, the performance of the proposed energy was dropped when the parameters A and D of the shape template were larger than 7 or 8 [pixel] which is roughly 10% of the object's diameter. The limit threshold values may depend not only on size of object but also on the shape and contrast of the boundary, parameter λ balancing the two costs in equation (2). Further experiments using various synthetic images and clinical data will be conducted in the near future to investigate the limit threshold values. Second, segmentation of flat lesions adherent to organ's surface, such as pleural plaque, remained unsolved. A novel energy will be developed and tested. Third, application to another segmentation problem, such as multi-organ segmentation and segmentation of 3D MR images, are of interest for future works.

References

1. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active Shape Models: Their Training and Application. CVIU 61(1), 38–59 (1995)
2. Lamecker, H., Lange, T., Seebass, M.: Segmentation of the Liver using a 3D Statistical Shape Model. Technical report ZIB-Report 04-09, Zuse Institue, Berlin (2004)
3. Leventon, M.E., Grimson, W.E.L., Faugeras, O.: Statistical Shape Influence in Geodesic Active Contours. In: CVPR, vol. 1, pp. 316–323 (2000)
4. Heimann, T.: Statistical Shape Models for 3D Medical Image Segmentation: A Review. Medical Image Analysis 13(4), 543–563 (2009)
5. Cremers, D., Rousson, M., Deriche, R.: A Review of Statistical Approaches to Level Set Segmentation: integrating Color, Texture, Motion and Shape. International Journal of Computer Vision 72(2), 195–215 (2007)
6. Heimann, T., Wolf, I., Meinzer, H.-P.: Active Shape Models for a Fully Automated 3D Segmentation of the Liver – An Evaluation on Clinical Data. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 41–48. Springer, Heidelberg (2006)
7. Okada, T., Shimada, R., Hori, M., Nakamoto, M., Chen, Y.W., Nakamaura, H., Sato, Y.: Automated Segmentation of the Liver from 3D CT Images using Probabilistic Atlas and Multi-level Statistical Shape Model. Academic Radiology 15(11), 1390–1403 (2008)
8. Boykov, Y., Funka-Lea, G.: Graph Cuts and Efficient N-D Image Segmentation. International Journal of Computer Vision 70(2), 109–131 (2006)
9. Slabaugh, G., Unal, G.: Graph Cuts Segmentation using an Elliptical Shape Prior. In: Proc. of ICIP, vol. 2, pp. 1222–1225 (2005)
10. Funka-Lea, G., Boykov, Y., Florin, C., Jolly, M., Moreau- Gobard, R., Ramaraj, R., Rinck, D.: Automatic Heart Isolation for CT Coronary Visualization using Graph-cuts. In: ISBI, pp. 614–617 (2006)
11. Veksler, O.: Star shape prior for graph-cut image segmentation. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 454–467. Springer, Heidelberg (2008)
12. Freedman, D., Zhang, T.: Interactive Graph Cut based Segmentation with Shape Priors. In: CVPR, vol. 1, pp. 755–762 (2005)

13. El-Zehiry, N., Elmaghraby, A.: Graph Cut Based Deformable Model with Statistical Shape Priors. In: ICPR, pp. 1–4 (2008)
14. Ali, A.M., Farag, A.A., El-Baz, A.S.: Graph cuts framework for kidney segmentation with prior shape constraints. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 384–392. Springer, Heidelberg (2007)
15. Park, H., Bland, P.H., Meyer, C.R.: Construction of an Abdominal Probabilistic Atlas and its Application in Segmentation. IEEE Trans. Med. Imag. 22(4), 483–492 (2003)

Author Index

Abugharbieh, Rafeef 204
Arbel, Tal 43

Baloch, Sajjad 11
Basso, Curzio 130
Bayouth, John 63
Beyerlein, Peter 21
Bhatia, Sudershan 63
Bi, Jinbo 118
Birngruber, Erich 86
Bischof, Horst 86
Boucher, Maxime 174
Brandt, Sami S. 1

Chen, Chao 31
Chiusano, Gabriele 130
Collins, D. Louis 43
Comaniciu, Dorin 96
Criminisi, Antonio 106

Davatzikos, Christos 164
de Bruijne, Marleen 153
Delgado Leyton, Edgar J.F. 141
Donner, René 86

Evans, Alan 174

Fang, Tong 11
Freedman, Daniel 31

Gangeh, Mehrdad J. 153
Gao, Yi 195

Hamarneh, Ghassan 204
Hamprecht, Fred A. 74
Han, Dongfeng 63
Hernández Hoyos, Marcela 141

Ishizu, Koich 214

Jannin, Pierre 54

Kamel, Mohamed S. 153
Kamen, Ali 164
Kaster, Frederik O. 74
Kelm, B. Michael 96
Khurd, Parmeshwar 164
Kikinis, Ron 195
Kobatake, Hidefumi 214
Konukoglu, Ender 106

Ladic, Lance 164
Lalys, Florent 54
Langs, Georg 86
Lorenz, Cristian 21
Lu, Le 118

Maday, Peter 164
Maes, Frederik 184
Menze, Bjoern H. 74
Morandi, Xavier 54

Nakagomi, Keita 214
Narihira, Takuya 214
Nawano, Shigeru 214
Nielsen, Mads 1

Orkisz, Maciej 141

Petersen, Kersten 1

Rajalingham, Rishi 43
Ribbens, Annemie 184
Riffaud, Laurent 54
Robertson, Duncan 106
Rose, Georg 21
Rosen, Mark 164
Ruppertshofen, Heike 21

Salah, Zein 21
Salganicoff, Marcos 118
Santoro, Matteo 130
Schmidt, Sarah 21
Schnall, Mitchell 164
Schramm, Hauke 21
Shaker, Saher B. 153
Shimizu, Akinobu 214
Shinozaki, Kenji 214
Shotton, Jamie 106
Siddiqi, Kaleem 174

Sørensen, Lauge 153
Sonka, Milan 63
Staglianò, Alessandra 130
Steiner, Helmut 86
Suehling, Michael 96
Suetens, Paul 184

Tannenbaum, Allen 195
Toews, Matthew 43
Togashi, Kaori 214
Top, Andrew 204

Vandermeulen, Dirk 184

Weber, Marc-André 74
Wels, Michael 96
Wolf, Matthias 118
Wu, Xiaodong 63

Zheng, Yefeng 96
Zhou, S. Kevin 96
Zouhar, Alexander 11
Zuluaga, Maria A. 141