Guang-Zhong Yang
Tianzi Jiang (Eds.)

LNCS 3150

Medical Imaging and Augmented Reality

Second International Workshop, MIAR 2004
Beijing, China, August 2004
Proceedings

Lecture Notes in Computer Science 3150

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

Editorial Board

This page intentionally left blank

Guang-Zhong Yang Tianzi Jiang (Eds.)

Medical Imaging and Augmented Reality

Second International Workshop, MIAR 2004
Beijing, China, August 19-20, 2004
Proceedings

Springer

eBook ISBN: 3-540-28626-8
Print ISBN: 3-540-22877-2

Created in the United States of America

Visit Springer's eBookstore at: http://ebooks.springerlink.com
and the Springer Global Website Online at: http://www.springeronline.com

Preface

Rapid technical advances in medical imaging, including its growing application to drug/gene therapy and invasive/interventional procedures, have attracted significant interest in close integration of research in life sciences, medicine, physical sciences and engineering. This is motivated by the clinical and basic science research requirement of obtaining more detailed physiological and pathological information about the body for establishing localized genesis and progression of diseases. Current research is also motivated by the fact that medical imaging is increasingly moving from a primarily diagnostic modality towards a therapeutic and interventional aid, driven by recent advances in minimal-access and robotic-assisted surgery.

It was our great pleasure to welcome the attendees to **MIAR 2004**, the 2nd International Workshop on Medical Imaging and Augmented Reality, held at the Xiangshan (Fragrant Hills) Hotel, Beijing, during August 19–20, 2004. The goal of **MIAR 2004** was to bring together researchers in computer vision, graphics, robotics, and medical imaging to present the state-of-the-art developments in this ever-growing research area. The meeting consisted of a single track of oral/poster presentations, with each session led by an invited lecture from our distinguished international faculty members. For **MIAR 2004,** we received 93 full submissions, which were subsequently reviewed by up to 5 reviewers, resulting in the acceptance of the 41 full papers included in this volume. For this workshop, we also included 4 papers from the invited speakers addressing the new advances in MRI, image segmentation for focal brain lesions, imaging support for minimally invasive procedures, and the future of robotic surgery.

Running such a workshop requires dedication, and we are grateful for the generous support from the Chinese Academy of Sciences. We appreciate the commitment of the MIAR 2004 Programme Committee and the 50 reviewers who worked to a very tight deadline in putting together this workshop. We would also like to thank the members of the local organizing committee, who worked so hard behind the scenes to make **MIAR 2004** a great success. In particular, we would like to thank Paramate Horkaew, Shuyu Li, Fang Qian, Meng Liang, and Yufeng Zang for their dedication to all aspects of the workshop organization.

In addition to attending the workshop, we trust that the attendees took the opportunity to explore the picturesque natural scenery surrounding the workshop venue. The Fragrant Hills Park was built in 1186 in the Jin Dynasty, and became a summer resort for imperial families during the Yuan, Ming and Qing Dynasties. We also hope some of you had the time to further explore other historical sites around Beijing including the Forbidden City, the Temple of Heaven, the Summer Palace and the Great Wall. For those unable to attend, we hope this volume will act as a valuable reference to the MIAR disciplines, and we look forward to meeting you at future MIAR workshops.

August 2004

Max Viergever, Xiaowei Tang,
Tianzi Jiang, and Guang-Zhong Yang

This page intentionally left blank

MIAR 2004
2nd International Workshop on Medical Imaging and Augmented Reality

General Co-chairs

Max Viergever, University Medical Center Utrecht, The Netherlands
Xiaowei Tang, Bio-X Laboratory, Zhejiang University, China

Executive General Chair

Tianzi Jiang, NLPR, Chinese Academy of Sciences, China

Program Committee Chair

Guang-Zhong Yang, Imperial College London, UK

Program Committee Members

Nicholas Ayache, INRIA, France
Hujun Bao, Zhejiang University, China
Wufan Chen, First Military Medical University, China
Ara Darzi, Imperial College London, UK
Brian Davies, Imperial College London, UK
David Firmin, Imperial College London, UK
David Hawkes, King's College London, UK
Karl Heinz Hoehne, University of Hamburg, Germany
Ron Kikinis, Brigham & Women's Hospital, Harvard, USA
Frithjof Kruggel, MPI for Cognitive Neuroscience, Germany
Qingming Luo, Huazhong University of Science and Technology, China
Shuqian Luo, Capital University of Medical Science, China
Xiaochuan Pan, University of Chicago, USA
Steve Riederer, Mayo Clinic, USA
Dinggang Shen, University of Pennsylvania School of Medicine, USA
Pengfei Shi, Shanghai Jiao Tong University, China
Jie Tian, Chinese Academy of Sciences, China
Yongmei Wang, Yale University School of Medicine, USA
Takami Yamaguchi, Tohoku University, Japan
Yan Zhuo, Institute of Biophysics, Chinese Academy of Sciences, China

Local Organization Co-chairs

Shuyu Li, NLPR, Chinese Academy of Sciences
Fang Qian, NLPR, Chinese Academy of Sciences

Organizing Committee Members

Yufeng Zang, NLPR, Chinese Academy of Sciences
Wanlin Zhu, NLPR, Chinese Academy of Sciences
Gaolang Gong, NLPR, Chinese Academy of Sciences
Meng Liang, NLPR, Chinese Academy of Sciences
Yong He, NLPR, Chinese Academy of Sciences
Longfei Cong, NLPR, Chinese Academy of Sciences
Chunyan Yin, NLPR, Chinese Academy of Sciences
Wei Zhao, NLPR, Chinese Academy of Sciences

Table of Contents

Invited Contributions

Image Processing, Reconstruction and Coding

Statistical and Shape Based Segmentation

Brain Image Analysis

Cardiac Modeling and Segmentation

Image Registration

Surgical Navigation and Augmented Reality

New Advances in MRI

Stephen J. Riederer, Ph.D.

MR Laboratory, Mayo Clinic, 200 First Street SW,
Rochester MN 55905 USA
riederer@mayo.edu

Abstract. Since its initial use in humans in the early 1980s magnetic resonance imaging (MRI) has become a widely used clinical imaging modality. Nonetheless, there continue to be opportunities for further advances. One of these is improved technology. Specific projects include high field strength magnets at 3 Tesla and beyond and an increased number of receiver channels for data acquisition, permitting improved SNR and reduced acquisition time. A second area is the further study of image formation, including the manner of sampling "k-space" and the specific type of image contrast. A third area is the in-creased exploitation of high speed computation to allow every-day implementation of techniques other-wise limited to research labs. Finally, MR is growing in its usage as a non-invasive, reproducible, and quantitative test in the study of non-clinical questions. MRI continues to be an area with a wealth of opportunity for contemporary study.

1 Introduction

Over the last two decades magnetic resonance imaging (MRI) has become a widely accepted technique useful for the clinical depiction of many types of pathologies of the brain, spine, abdomen, and musculoskeletal and cardiovascular systems. The significance and impact of this can be seen in various ways. For example, currently there are approximately 15,000 whole body MRI units installed worldwide with approximately 7,000 of these installed in the United States [1]. With a very conservative estimate of ten clinical examinations per scanner per day, this converts to well over 100,000 MRI exams daily around the world. Another measure is the continuing growth of clinical MRI. Although there have been year-to-year variations, over the ten-year period from 1992 to 2002 MRI at Mayo Clinic grew at a 10.4% annual rate, and this is typical of many institutions. Yet another measure of the significance of the modality was the awarding of the 2003 Nobel Prize in the category of Physiology or Medicine to two pioneers in MRI development, Drs. Paul Lauterbur and Peter Mansfield. By each of these measures MRI has become significant in modern medicine around the world.

In spite of this success and clinical acceptance there is still ample room for MRI to grow technically and scientifically. The fundamental limitations of MRI, primarily the limits in the acquisition speed and the signal-to-noise ratio (SNR), have still not been adequately addressed in many applications. Also, the fundamental technical

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 1-9, 2004.

advantages of MRI over other modalities, such as the high degree of contrast flexibility and the arbitrary nature in which an image can be formed, can be further exploited.

The purpose of this work is to describe several contemporary trends in the ongoing technical development of MRI. These include advances in technology for providing improved MRI data, advances in the manner of sampling MRI data acquisition space or "k-space," computational techniques for facilitating the high-speed formation of MR images, and advances in the manner in which MRI is used as a scientific tool.

2 MRI Technology

2.1 Increased Magnet Strength

Selection of field strength is one of the most important choices in defining an MRI system, as it drives many of the other aspects of the system, such as siting, available contrast by virtue of the field-dependent relaxation times, and intrinsic SNR. In the mid-1980s MRI manufacturers developed systems at a field strength of 1.5 Tesla, and this became the *de facto* maximum operating strength which was routinely available. Since that time a number of applications have been identified as potentially benefiting from higher strength, such as *in vivo* MR spectroscopy, functional neuro MRI using BOLD contrast, and SNR-starved applications such as those using various fast-scan techniques. The advantages of higher field are offset by increased specific absorption rate (SAR) and decreased penetration of the RF field into the body. To address this interest, MR vendors in approximately the last five years have developed systems at 3.0 Tesla for routine installation. Additionally, whole body systems have been developed for individual research laboratories at 4, 7, 8, and 9 Tesla. The advantages of such systems in the applications indicated above are in the process of being studied. It remains to be seen to what extent these advantages trigger widespread installation.

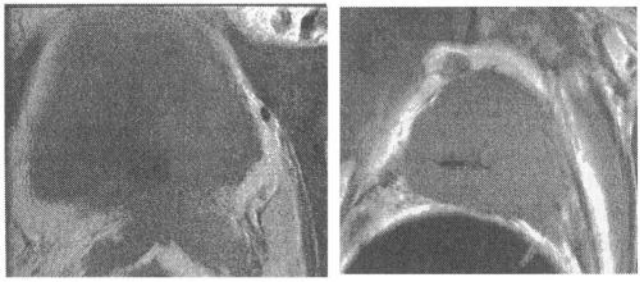


Fig. 1. Comparison of image of prostate at 1.5 Tesla (a, left) and 3.0 Tesla (right). Note the improved SNR of the latter due to the increased field strength. The same T1-weighted spin-echo sequence was used for both

An example of the advantages of 3.0 Tesla is shown in Figure 1. Figure 1a is an image of the prostate of a human cadaver acquired at a field strength of 1.5 Tesla. A four-element surface receiver coil was used in conjunction with a 12 cm FOV and

256x256 spin-echo imaging sequence. Fig. 1b is an image of the same specimen using the same pulse sequence and scan parameters as for (a) but now at a field strength of 3.0 Tesla. The improvement in SNR is readily apparent, measured in this study as 2.1X. Results such as these may increase the clinical applicability of MRI in these anatomic regions as well as in other areas in which SNR can be limiting. A specific example is high resolution imaging of the hand. In these cases it is critical that appropriate receiver coils be used which are not only tuned to the increased resonant frequency but also matched to the anatomic region under study.

2.2 Improved Receiver Channels

Another area of technology in which there has been considerable recent development is in the receiver chain of the MRI scanner. Receiver coils have long been a field of study in MRI. Early developments included developing a single "coil" consisting of several distinct coil elements, the signals from which were added prior to digitization and reconstruction [2]. In ca. 1990 multi-coil arrays were developed in which a separate image was made from each individual coil element, the results then added in quadrature to improve SNR [3]. With this approach the signal from each receiver element was directed to an individual receiver channel, and typically the number of such channels was limited to four. However, recently there has been interest in expanding the number of such receiver channels, as motivated by the desire for further gains in SNR, broader anatomic coverage, and the implementation of various parallel imaging techniques. Thus, modern, top-of-the-line scanners are equipped with 8, 16, and even 32 individual receiver channels. Additional flexibility is provided by allowing for coil arrays with even more elements than the number of channels. Switching is allowed to direct a specific element to a specific receiver.

Figure 2 is a comparison of a coronal scan of the abdomen using a single-shot fast-spin-echo technique, the result in (a) acquired using a standard four-element phased array, that in (b) formed using a modern eight-element coil with eight receiver channels. The pulse sequence was identical for the two scans. The result in (b) is clearly superior in SNR.

3 MR Image Formation

3.1 k-space Sampling Techniques

The measured signal in MRI samples the Fourier transform of the final image. This is often referred to as "k-space." Early theory showed that the time integral of the grad-

ent waveforms was proportional to the specific k-space position being sampled. Thus, customized gradient manipulations could provide very precise k-space trajectories. Over time, the most commonly used method has been a 2DFT technique in which an individual line is sampled each repetition of the acquisition, the final image formed from a data set comprised from a set of such parallel lines. However, other trajectories have also been developed in the last two decades, most notably radial or projection reconstruction (PR) and spiral. Each has its specific advantages and limitations. However, recently further variants in these have been developed, in some cases allowing for time-resolved imaging.

One example of a such a technique is time-resolved imaging of contrast kinetics or "TRICKS" [4]. This is a 3D acquisition method in which the central portion of k-space is sampled periodically, and outer regions are sampled less frequently. The final image set is reconstructed from the most recent measurement of each phase encoding view. Because of the difference in sampling rates, the actual rate at which images are formed is greater than that which is dictated by the spatial resolution: Ny x Nz x TR.

Other means for k-space sampling have also been developed. One recently described technique is termed "PROPELLER" [5] because of the manner in which a set of vanes is sampled, similar to those comprising a propeller or windmill. The width of each vane may consist of ten or more individual phase encoding views. Because each vane intersects the region in the immediate vicinity of the k-space origin, the redundancy of information potentially allows some immunity to motion artifact as well as the ability to generate a time-resolved image sequence.

Another MR data acquisition technique recently described combines the view sharing of TRICKS, the view ordering of elliptical centric (EC) acquisition [6], and the radial sampling of PR and PROPELLER techniques and uses a star-like pattern to generate a time series of 3D images. The EC-STAR pattern is shown in Figure 3. Each point in the figure corresponds to an individual measurement as sampled in a single repetition of the 3D acquisition. As shown, k-space is divided into three distinct regions: the central disk (R1) and two annular rings (R2 and R3). These regions are sampled at the relative rates of 1, ½, and ¼, respectively. The time series thereby generated is roughly three times the frequency intrinsic to the sampling. Also, by sampling all of the central views in a group the technique has improved immunity to artifact and reduced latency. This technique has recently been applied to MR imaging in conjunction with continuous table motion [7].

3.2 Parallel Image Reconstruction Techniques

Recently a number of techniques have been described in which the redundancy of information provided from multiple coil elements can be used to reduce the acquisition time for image formation. The two general classes of methods are referred to as "SMASH" [8] and "SENSE" [9]. Here we briefly describe the latter which has thus far been implemented to a wider extent than the former.

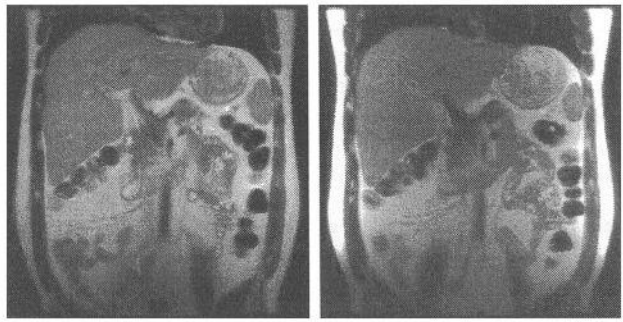


Fig. 2. Comparison of image of the abdomen formed using a standard four-element phase array coil (left) and an eight-element coil (b, right) using the same breath-hold T1-weighted pulse sequence. The result in (b) has improved SNR

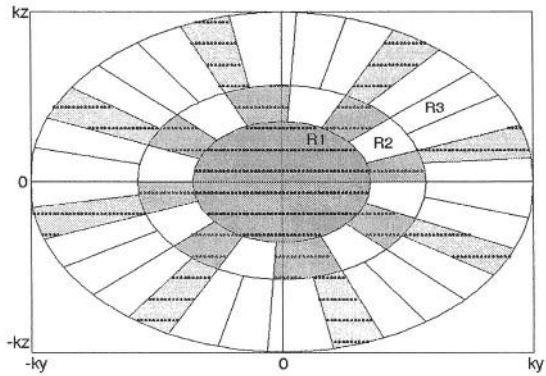


Fig. 3. Sampling pattern in ky-kz space for elliptical centric sampling with a star-like trajectory (EC-STAR). The relative sampling frequencies are unity, one-half, and one-fourth for the central disk and the two annual rings. The technique can be used to generate time-resolved 3D data sets of contrast agents flowing through the vasculature

The concept behind SENSE is to allow aliasing in the MR image by using an acquisition field-of-view (FOV) which is smaller than the actual size of the object. In the resultant image this causes the edges of the object to be aliased into the central portion. If uncorrected, this often causes the final MR image to be uninterpretable. The basis of SENSE is to use multiple receiver coil elements and generate a separate aliased image for each element. If the coil sensitivities are known and distinct from each other, this information can be used to determine the constituent signals comprising the aliased image and essentially restore the full field of view with no aliasing. This technique can be used in various ways, the two principal ones being: (i) to reduce the acquisition time for a given spatial resolution, and (ii) to improve the spatial resolution for a given acquisition time. The SENSE technique can be useful in applications in which it is important to acquire all of the MR data within a limited time, such as a breathhold.

An example is shown in Figure 4. Figure 4a is a contrast-enhanced MR angiogram in coronal orientation at the level just below the knees. The acquisition time was 30 sec for the 3D sequence. Figure 4b used the same pulse sequence and the same acquisition time except that SENSE encoding was performed. This allowed an improvement in spatial resolution along the left-right direction by a factor of two. The resultant improved sharpness in the image in (b) is apparent vs. (a).

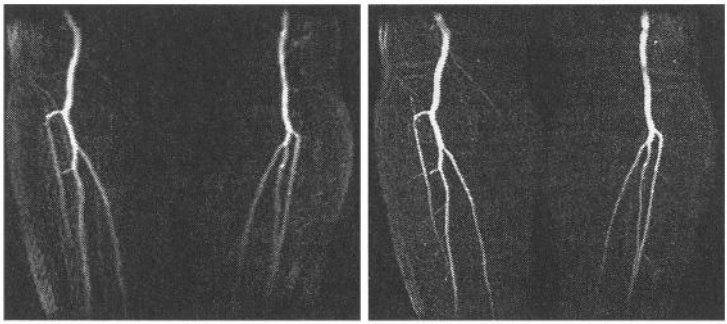


Fig. 4. Contrast-enhanced MR angiograms of the lower legs acquired using a 30-second long 3DFT acquisition. The image in (a,left) used a standard 3DFT acquisition. The image in (b, right) was formed using SENSE with a two-fold improvement in resolution. Improved sharpness in the left-right direction of (b) vs. (a) is apparent

4 Computational in MRI

MRI is a very computationally intensive technique. Because the MRI data are acquired in k-space, all images are formed only after some kind of image reconstruction process, typically done using Fourier transformation. In the 1980s the image formation process in MRI typically took tens of minutes. Data collection was time consuming as was the image reconstruction. However, as scan times have dropped in the last two decades so too has the demand for faster image reconstruction increased. This has been addressed to some extent by the ever-increasing computational speed of computers, and today for a variety of pulse sequences the MR images can be reconstructed in real time with the data acquisition.

Increased computational speed is also important as the number of receiver channels increases. As discussed earlier, this was motivated by the desire for improved SNR, but a consequence of now allowing N receiver channels is that there is N times as much data and N times as many image reconstructions to perform compared to a reference acquisition. Improved computational hardware can potentially address this.

Another application of improved computational speed is the implementation of more advanced image correction or reconstruction methods which might otherwise not be practical. One example is the use of "gridding" techniques [10] which perform the interpolation onto a rectilinear grid of data which may have been acquired along some non-rectilinear k-space trajectory. This is critical in spiral acquisition methods.

A very specific example of data correction is the process of accounting for gradient warping in the imaging of a long field of view during continuous motion of the patient table. The motivation for this project is to match the velocity of the table to the velocity of the contrast bolus as it flows from the thorax to the abdomen and the legs. This is critical in the performance of contrast-enhanced angiograms of the peripheral vasculature. It is well known in imaging using a conventional static patient table that non-linearities in the magnetic gradients cause distortion in the reconstructed image. Specifically, the coronal-oriented image of a square grid typically is distorted to resemble a barrel. In virtually all modern MRI systems this artifact is somehow accounted for. However, in continuous table motion MRI the problem is exacerbated because individual MR measurements are subjected to different degrees of distortion as a consequence of their motion through the gradient field.

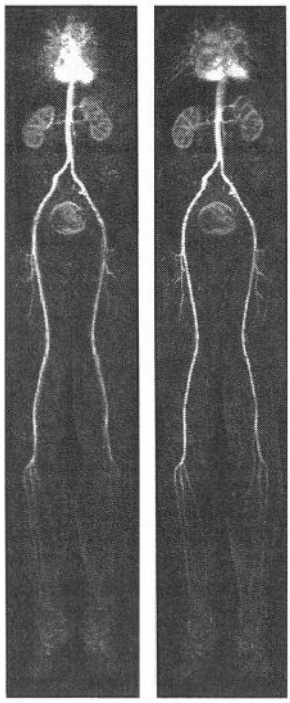


Fig. 5. Comparison of contrast-enhanced MR angiograms of the abdomen, pelvix, and legs of the same data set uncorrected (a, left) and corrected (b, right) for gradient non-linearity in moving table MRI. Note the improved sharpness of the small vessels in the thighs and calves in the corrected image

The above problem can be accounted for but it is computationally intensive. Polzin and colleagues have described a method [11] in which each individual phase encoding view is reconstructed into an image component, the component is then corrected for the gradient warping, and then the corrected components are added together to form the final 3D data set. This is time consuming because rather than perform one 3D Fourier transform on the complete data set once it is acquired, in essence a separate 3D Fourier transform must be performed for each individual measurement. In practice this can be relaxed somewhat in that the transform can be done in groups of approximately several dozen views. Nonetheless, without today's computational power such algorithms would not be practical for every day use. An example of gradient warping correction applied to moving table MRI is shown in Figure 5.

5 MRI as a Quantitative Test

For many of the same reasons that MRI has become clinically accepted it is also becoming widely used as a reference standard in the testing of various scientific hypotheses. Specifically, MRI is quantitative, non-invasive, reproducible, three-dimensional, and non-destructive, it uses no ionizing radiation, and it has a high degree of contrast flexibility. It is also generally affordable.

One prime example of how MRI is used in this fashion is in the radical increase in the amount of research being conducted in studies of the brain using functional neuro MRI with the BOLD response. Although there continue to be technical advances in the manner in which BOLD images are acquired and processed, the basic mechanism is relatively well understood, has been widely implemented, and is being widely used by investigators worldwide in many of the various neurosciences.

Another example of the manner in which MRI is used as a scientific standard is for many kinds of animal studies. This can be used to assess phenotype, follow disease progression, and monitor the effectiveness of therapy or drug response. Individual animals can be used in longitudinal studies, obviating the need to sacrifice a portion of the cohort at each phase. Special purpose, small bore, high field MRI systems are becoming more common to facilitate such investigations.

The contrast flexibility of MRI is another potential factor contributing to its increased use. One emerging type of contrast is to use the MRI signal to measure the elasticity of materials using the method dubbed "MR elastography" [12]. With this technique the manner in which acoustic waves propagate through the medium is used to estimate the wave velocity and material stiffness or elastic modulus.

The use of MRI as a quantitative test will no doubt grow in the near future as it becomes more turnkey, more accessible, and as its advantages are more widely appreciated.

6 Summary

Although it has been widely accepted clinically for over two decades, MRI is still undergoing considerable technical development. This includes MRI technology itself, the specific means for acquiring MR image data in k-space and reconstructing the image set, computational hardware allowing sophisticated image formation and correction algorithms, and special purpose scanners to facilitate the utilization of MRI as a scientific standard.

Acknowledgments

The author gratefully acknowledges the contributions of Joel P. Felmlee, James F. Glockner, Houchun H. Hu, David G. Kruger, Ananth J. Madhuranthakam, Jason A. Polzin, and David W. Stanley.

References

1. Frost and Sullivan Marketing Agency. USA MRI Market. (2002)
2. Bruce, K.B., Cardelli, L., Pierce, B.C.: Comparing Object Encodings. In: Abadi, M., Ito, T. (eds.): Theoretical Aspects of Computer Software. Lecture Notes in Computer Science, Vol. 1281. Springer-Verlag, Berlin Heidelberg New York (1997) 415–438
3. Roemer, P.B., Edelstein, W.A., Hayes, C.E., Souza, S.P., Mueller, O.M.: The NMR phased array. Magn. Reson. Med. 16 (1990) 192-225.
4. Korosec, F.R., Frayne, R., Grist, T.M., Mistretta, C.A.: Time-resolved contrast-enhanced 3D MR angiography. Magn. Reson. Med. 36 (1996) 345-351.
5. Pipe, J.G.: Motion correction with PROPELLER MRI: application to head motion and free-breathing cardiac imaging. Magn. Reson. in Med. 42 (1999) 963-969.
6. Wilman, A.H., Riederer, S.J., King, B.F., Debbins, J.P., Rossman, P.J., Ehman, R.L.: Fluoroscopically-triggered contrast-enhanced three-dimensional MR angiography with elliptical centric view order: application to the renal arteries. Radiology 205 (1997) 137-146.
7. Madhuranthakam, A.J., Kruger, D.G., Riederer, S.J., Glockner, J.F., Hu, H.H.: Time-resolved 3D contrast-enhanced MRA of an extended FOV using continuous table motion. Magn Reson in Med 51 (2004) 568-576.
8. Sodickson, D.K., Manning, W.J.: Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays. Magn. Reson. Med. 38 (1997) 591-603.
9. Pruessman, K.P., Weiger, M., Scheidegger, M.B., Boesiger, P.: SENSE: sensitivity encoding for fast MRI. Magn. Reson. Med. 42 (1999) 952-962.
10. O'Sullivan, J.D.: A fast sinc function gridding algorithm for Fourier inversion in computer tomography. IEEE Trans. Med. Img. MI-4 (1985) 200-207.
11. Polzin, J.A., Brittain, J.H., Gur, D.H., Debbins, J.P., Kruger, D.G., Riederer, S.J.: Correction for gradient field distortions in moving table imaging. Pres. at 10th Mtg, ISMRM, Honolulu, (2002) 380.
12. Muthupillai, R., Lomas, D.J., Rossman, P.J., Greenleaf, J.G., Manduca, A., Ehman, R.L.: Magnetic resonance elastography by direct visualization of propagating acoustic strain waves. Science 269 (1995) 1854-1857.

Segmentation of Focal Brain Lesions

Frithjof Kruggel

Interdisziplinäres Zentrum für Klinische Forschung (IZKF),
Inselstrasse 22, D-04103 Leipzig, Germany
kruggel@cbs.mpg.de

Abstract. Focal brain lesions are a consequence of head trauma, cerebral infarcts or intracerebral hemorrhages. In clinical practice, magnetic resonance imaging (MRI) is commonly used to reveal them. The segmentation task consists of finding the lesion borders. This problem is non-trivial because the lesion may be connected to other intracranial compartments with similar intensities. A new method for the automatic segmentation of unilateral lesions is proposed here. The signal statistics of multichannel MR are examined w.r.t. the first-order mirror symmetry of the brain. The algorithm is discussed in detail, and its properties are evaluated on synthetic and real MRI data.

1 Introduction

High resolution magnetic resonance (MR) images of the brain are used in clinical practice to reveal focal brain lesions (e.g., as consequences of head trauma, intra-cerebral hemorrhages or cerebral infarcts). Lesion properties (i.e., position, extent, density) are known to be related to cognitive handicaps of a patient. While a *semi-quantitative* analysis of MR tomograms based on visual inspection (e.g., rating scales) is common today in certain clinical protocols, tools for a *quantitative* analysis are still rare. One of the reasons for this lack of tools is that segmenting MR images with pathological findings is considered a non-trivial task.

Manual lesion segmentation is still considered as the "gold standard". A human expert with anatomical knowledge, experience and patience uses some graphical software tool to outline the region of interest. While this method obviously produces the most reliable results, it is time consuming and tedious. In addition, re-tests and inter-rater reliability studies of manually segmented lesion rarely reach 90 % correspondence [2], [21]. Most previous studies in automatical lesion segmentation concentrated on the detection of white matter lesions in Multiple Sclerosis (MS). Techniques suggested for this problem include: statistical clustering [19], a combination of statistical techniques and anatomical knowledge [7], a combined classification of multi-channel MR images [22] or an iterative approach to correct B_1 field inhomogeneities while classifying voxels [11]. However, the problem studied in this paper is more general. While MS lesions are completely caused by white matter, lesions as consequences of a head trauma or cerebral infarction may include the cortical gray matter and thus reach the cerebrospinal fluid (CSF) compartment. So the problem is to discriminate a lesion from *different* surrounding compartments.

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 10–18, 2004.

Few semi-automatic and automatic methods exist in the literature for this problem. Most are dedicated to the segmentation of a specific type of focal lesion only. Maksimovic *et al.* [14] studied the segmentation of fresh hemorrhagic lesions from CT data using 2D active contours. Such lesions have high signal intensities and may reach the skull which is also bright. The active contour detects the border between the brain and the skull, the boundary of the ventricles and the boundary of the lesion. Loncaric *et al.* [12], [13] proposed an approach that combines unsupervised fuzzy clustering and rule-based system labeling. The rule-based system assigns one of the following labels to each region provided by the clustering: background, skull, brain, calcifications and intracerebral hemorrhages.

Dastidar *et al.* [3] introduced a semi-automatic approach for segmenting infarct lesions consisting of four steps: image enhancement, intensity thresholding, region growing and decision trees in order to localize the lesion. User interaction is required to define the lesion boundaries if it reaches a compartment of similar intensity. Stocker *et al.* [17] proposed to automatically classify multichannel image information (T_1-, T_2- and proton density (PD)) using a a self-organizing map into five partitions: white and gray matter, CSF, fluid and gray matter in the infarct region. Brain tumors may be segmented using statistical classification [8], [15]. An atlas of normal brain anatomy containing spatial tissue probability information is used to discriminate different anatomical structures with similar intensities. A tumor is found as a (compact) region of outlier voxels.

A level-set method guided by a tumor probability map was described by Ho *et al.* [4]. Finally, a region growing technique was proposed to segment any type of lesions [18]. It requires the input of a seed point and a pre-defined threshold to avoid an overgrowing outside the lesion. A similar method was developed by Hojjatoleslami *et al.* [5], [6]. The key idea is to stop the region growing on the outer cortical layer between the lesion and the external CSF area, that is often preserved after stroke. The algorithm involves a grey level similarity criterion to expand the region and a size criterion to prevent from overgrowing outside the lesion.

In this paper, we focus on the segmentation of unilateral focal brain lesions in their chronic stage. Lesions are generally not homogeneous, often with completely damaged core parts and minor damage in peripheral portions. Thus, MR signal intensities range between values of undamaged tissue and values similar to CSF. The boundary between a cortical lesion and the CSF compartment is often hard to draw.

The following section of this paper describes the method. In a subsequent section, we study the parameter settings and performance of our method by several exeriments. Finally, properties of this approach are summarized.

2 The Algorithm

As a first approximation, the brain is a mirror-symmetric organ. Lesions considered here are confined to a single hemisphere with a generally healthy area on the contralateral side (see Fig. 1. The segmentation problem may therefore be stated as finding compact areas with an intensity statistic that differs significantly from the contralateral side. A Hotelling T^2 test is performed to compare small subregions from both hemispheres. The test measure is converted into a z-score and collected in a lesion probability map (LPM).

Areas with high signal asymmetries are depicted by high z-scores. A post-processing step thresholds the LPM and checks the size of all connected regions against the size distribution of natural asymmetries.

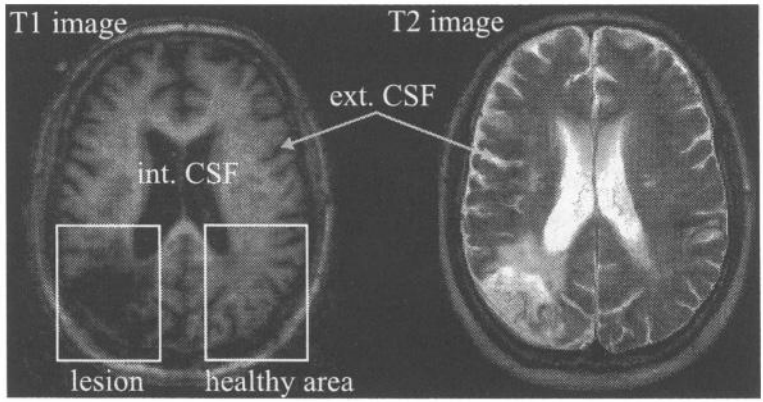


Fig. 1. Lesions considered here are confined to a single hemisphere with a generally healthy area on the corresponding contralateral side

2.1 Computing a Lesion Probability Map

To detect a probable lesion, we compare the signal statistics in homologeous regions of both brain hemispheres in multichannel MRI tomograms. As a pre-processing step, we align brain datasets with the stereotactical coordinate system such that the midsagittal plane is at a known location $x_{mid} = const.$ We use a natural convention for addressing the body side, i.e. locations $x_l < x_{mid}$ refer to the left body side. Now consider a cubic subregion R in the left brain hemisphere centered around a voxel v_l at Cartesian coordinates (x_l, y, z) with an extent of s voxels. Its homologeous region is centered around voxel v_r at $(2 * x_{mid} - x_l, y, z)$. At each voxel v, a vector of observed signal intensities $\mathbf{o}_k = \{o_1, \ldots, o_k\}$ is obtained from the k multichannel images. Thus, a region includes $n = s^3$ observation vectors.

We are now interested whether the multivariate mean of observations in both regions is different. Hotelling's T^2 statistic is an important tool for inference about the center of a multivariate normal quantity. According to Rencher [16], we can work directly with the differences of the paired observations from both hemispheres, i.e. reduce the problem to a one-sample test for $D_n = \{\mathbf{d_1}, \ldots, \mathbf{d_n}\}$ from $N_k(\bar{\mathbf{d}}, \mathbf{\Sigma})$, where $\mathbf{d_i} = \mathbf{o_{l,i}} - \mathbf{o_{r,i}}$ correspond to the left-right differences, and $\bar{\mathbf{d}}, \mathbf{\Sigma}$ are unknown. The hypothesis $H_0 : \bar{\mathbf{d}} = 0$ is rejected at the level α if

$$T^2 := n\bar{\mathbf{d}}^{\mathrm{T}}\mathbf{S}^{-1}\bar{\mathbf{d}} > \frac{(n-1)k}{(n-k)} F_{k,n-k,1-\alpha}, \tag{1}$$

where

$$\bar{\mathbf{d}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{d}_i \quad \text{and} \quad \mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{d_i} - \bar{\mathbf{d}})(\mathbf{d_i} - \bar{\mathbf{d}})^{\mathsf{T}} \tag{2}$$

are the mean and the covariance of the sample. Obtained $F_{k,q}$-scores are converted into significance levels p [1]:

$$p \in [0,1] = 2\,I_x\left(\frac{k}{2}, \frac{q}{2}\right), \quad \text{where} \quad x = \frac{q}{q + kF}, \tag{3}$$

and $I_x(\cdot)$ corresponds to the incomplete beta function [1]. Significance levels are finally converted into z-scores:

$$z = \sqrt{2}\,\mathrm{erfc}^{-1}(p/2), \quad p \in [0,1]. \tag{4}$$

Z-scores are compiled in a statistical image that we denote as LPM. Note that this map is symmetric with respect to the midsagittal plane.

2.2 A Weighted Hotelling T^2 Test

In order to obtain a more localized LPM, we include weights w_i with differences $\mathbf{d_i}$. The highest weight is addressed to the center voxel v_c of the subregion R. Weights decrease with distance from this voxel. A reasonable choice is a Gaussian weighting function:

$$w_i = \exp\left(-\frac{||v_i - v_c||^2}{2\sigma^2}\right), \quad v_i, v_c \in R, \tag{5}$$

where $||v_i - v_c||$ denotes the Euclidean distance between v_i and v_c and σ is a spatial scaling factor. Note that $\sigma \to \infty$ approaches the unweighted case above. Now, the weighted sample mean and covariance are computed as:

$$\bar{\mathbf{d}}_{\mathbf{w}} = \frac{\sum_{i=1}^{n} w_i \mathbf{d}_i}{\sum_{i=1}^{n} w_i} \quad \text{and} \quad \mathbf{S}_{\mathbf{w}} = \frac{\sum_{i=1}^{n} w_i(\mathbf{d_i} - \bar{\mathbf{d}})(\mathbf{d_i} - \bar{\mathbf{d}})^{\mathsf{T}}}{\sum_{i=1}^{n} w_i - \sum_{i=1}^{n} w_i^2}. \tag{6}$$

As Willems *et al.* [20] discussed for the case of a robust (weighted) Hotelling test, the test statistic is now:

$$T_w^2 := n\bar{\mathbf{d}}_{\mathbf{w}}^{\mathsf{T}}\mathbf{S}_{\mathbf{w}}^{-1}\bar{\mathbf{d}}_{\mathbf{w}} \approx fF_{k,q,1-\alpha}, \tag{7}$$

where f is a multiplication factor and q the modified degrees of freedom for the denominator of the F-distribution, given by:

$$f = E[T_w^2]\frac{q}{q-2} \quad \text{and} \quad q = 4 + \frac{2E^2[T_w^2](k+2)}{kVar[T_w^2] - 2E^2[T_w^2]}. \tag{8}$$

Since the mean and the variance of the T_w^2 distribution cannot be obtained analytically, we determined values of $E[T_w^2]$ and $Var[T_w^2]$ using Monte-Carlo simulations [20].

For a fixed dimension $k = 2$, we generated $m = 10^6$ samples $X_i; i = 1, \ldots, m$ from a $N_k(0, \mathbf{I}_k)$ Gaussian distribution. For each sample, $T_w^{2(i)}$ was determined by Eq.

7, using different region extents of $s = \{3, 5, 7\}$ voxels (corresponding to samples sizes of $n = \{27, 125, 343\}$) and different spatial scaling factors $\sigma = \{0.5, \ldots, 2.5\}$. The mean and variance of $T_w^{2(i)}$ are given by:

$$\hat{E}(T_w^2) := \frac{1}{m} \sum_{i=1}^{m} T_w^{2(i)} \quad \text{and} \quad \widehat{Var}(T_w^2) := \frac{1}{m-1} \sum_{i=1}^{m} (T_w^{2(i)} - \hat{E}(T_w^2))^2. \tag{9}$$

Following [20], a smooth function was fit to a regression model, depending on the window size s and the spatial scaling factor σ. For the reduced degrees-of-freedom q, given $k = 2$, we modelled:

$$q = 4 + \frac{4}{t_3 - 1}, \tag{10}$$

and, likewise, for the multiplication factor f:

$$f = (t_2 + t_1 \sigma^{-t_0}) \frac{q}{q-2}. \tag{11}$$

Values for t_i are given in Tab. 1.

s	t_0	t_1	t_2	t_3
3	20.74224	0.481705	2.347796	1.17270
5	4.02456	4.242255	2.066018	1.00784
7	3.78378	14.393344	2.006746	1.01824

Table 1. Regression parameters t_i for computing the multiplication factor f and the reduced degrees-of-freedom q, given a window size s.

The lesion probability map (LPM) is thresholded by z_{lim}, and the size of the connected components is determined. Natural asymmetries occur in any brain, but they are generally small compared with a brain lesion. The distribution of "pseudo-lesions" due to brain asymmetry was sampled from 20 datasets of healthy subjects. The size of a probable lesion is compared with this distribution, and a p-value is addressed to each lesion for being a "true" lesion. The algorithm was implemented and evaluated using the BRIAN system [9].

3 Experiments

The first experiment was conducted in order to study the influence of the parameters window size s, spatial scaling factor σ, and z-score threshold z_{lim} on the estimated size of the detected lesion. Denote a contrast ratio of 1.0 as a complete lesion, 0 as undamaged tissue. Simulated datasets with a lesion size of $l = \{13^3, 27^3, 58^3\}$ voxels and a contrast ratio of $c = \{0.1, 0.8\}$ were generated. The lesion detection algorithm was run on these data using window sizes of $s = \{3, 5, 7\}$, and $\sigma = \{0.5, 1.0, 1.5, 2.0, 2.5\}$.

We found: (1) The larger the window size, the higher the z-scores. (2) The larger σ, the higher the z-scores. (3) The z-score range, in which the true lesion size was correctly estimated, decreases with increasing σ. (4) The z-score range, in which the true lesion size was correctly estimated, increases with increasing window size. The best compromise was found with $s = 5$, $\sigma = 1$, and $z_{lim} = 4.3$ (see Fig. 2).

As a second experiment, we examined the contrast ratio, for which at least 95% of the true lesion size was detected. For a small lesion (10^3 voxels), $c = 0.17$ (3% noise), $c = 0.28$ (6% noise), for a large lesion (30^3 voxels), $c = 0.09$ (3% noise), $c = 0.18$ (6% noise) was found. So, for realistic noise levels found in MRI datasets, lesions with a contrast ratio of at least 0.2 are expected to be detected with a size that is close to the real one.

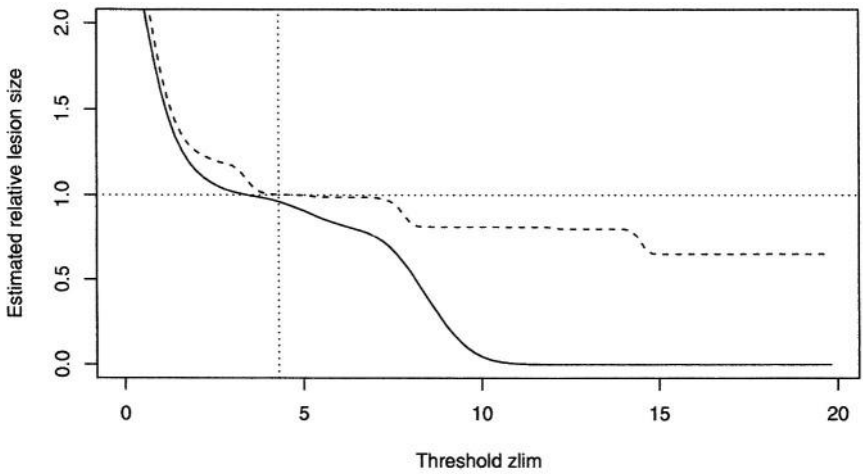


Fig. 2. Estimated relative size of the lesion vs. threshold z_{lim} for window size $w = 5$ and spatial scaling factor $\sigma = 1.0$. The solid line corresponds to a lesion contrast of 0.1, the broken line to a lesion contrast of 0.8.

Then, we were interested in discriminating real lesions from pseudo-lesions due to natural brain asymmetry which are expected to be small. We selected 20 datasets of healthy subjects from our brain database. A MDEFT protocol [10] was used to acquire high-resolution T_1-weighted data sets on a 3.0 Tesla Bruker Medspec 100 system (128 sagittal slices of 256*256 voxels, FOV 250 mm, slice thickness 1.4 mm, subsequent trilinear interpolation to an isotropic resolution of 1 mm). T_2-weighted datasets were collected on the same scanner (20 slices of 256*256 voxels of 0.97*0.97*7 mm). The T_1-weighted dataset was aligned with the stereotactical coordinate system, and the T_2-weighted dataset was rigidly registered with the aligned dataset using a multi-resolution approach and a cost function based on normalized mutual information. Then, the lesion segmentation algorithm was applied to the multichannel image, and the size of the de-

tected pseudo-lesions determined using $z_{lim} = 10$. In total, 2077 regions were found, and from their cumulative distribution function, a lesion of more than 880 voxels may be called pathological with an error of 5%.

Finally, we illustrate the use of this algorithm in a real dataset. A patient suffering from a stroke in the left anterior area of the middle cerebral artery was examined 6 months post-stroke (see Fig. 3). Note that not only the lesion itself is detected but also other areas (i.e., the left ventricle) are marked where some substance loss occured in the vicinity. Thus, all consequences of the stroke are depicted. Note further that low-intense regions in the vicinity of the Sylvian fissure are not included in the lesion, because they are symmetric.

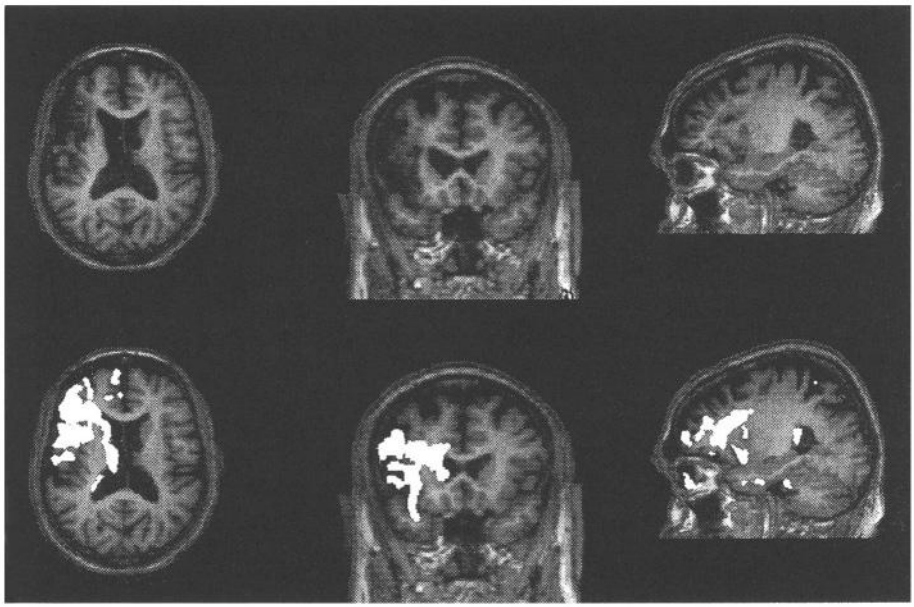


Fig. 3. Top: T_1-weighted image of a patient suffering from a cerebral infarction in the anterior supply area of the middle cerebral artery. Below: segmented lesion as detected by this algorithm.

4 Summary

We described an algorithm for detecting unilaterial focal lesions in MR images of the human brain. The signal statistic of small mirror-symmetric subregions from both hemispheres is compared using a spatially weighted Hotelling T^2 test. The resulting voxel-wise test measure is converted to a z-score and collected in a lesion probability map. This map is thresholded by a pre-determined z-score limit, and the size of the connected lesion components is computed. A lesion is detected by this algorithm with a size error

of less than 5% if the contrast ratio is at least 0.2. It may be denoted a "true lesion" with an error probability of 5% if it is bigger than 880 voxels. Currently, we analyze temporal changes of incompletely damaged tissue in a longitudinal study.

Acknowledgements

The author wishes to thank the MPI of Human and Cognitive Brain Science, Leipzig, for providing the datasets. The help of Dr. Claire Chalopin in the development of this algorithm during her visit at MPI is greatfully acknowledged.

References

1. Abramowitz, M., Stegun, I.A. (1992) Handbook of mathematical functions with formulas, graphs, and mathematical tables. United States Department of Commerce, Washington.
2. Bello, F., Colchester, A.C.F. (1998) Measuring global and local spatial correspondence using information theory. In: Proc. MICCAI 1998, Lecture Notes in Computer Sciences **1496,** 964–973.
3. Dastidar, P., Heinonen, T., Ahonen, J., Jehkonen, M. Molnar, G. (2000) Volumetric measurements of right cerebral hemisphere infarction: use of a semiautomatic MRI segmentation technique. Comp. Biol. Med. **30**, 41–54.
4. Ho, S., Bullitt, E., Gerig, G. (2002) Level-set evolution with region competition: automatic 3D segmentation of brain tumors. In: Proc. ICPR 2002, 532–535.
5. Hojjatoleslami, S.A., Kittler, J. (1998) Region growing: A new approach. IEEE Trans. Image Proc. **7,** 1079–1083.
6. Hojjatoleslami, S.A., Kruggel, F. (2001) Segmentation of large brain lesions. IEEE Trans. Med. Imag. **20**, 666–669.
7. Kamber, M., Shinghal, R., Collins, L., Francis, G.S., Evans, A.C. (1995) Model based 3D segmentation of multiple sclerosis lesions in magnetic resonance brain images. IEEE Trans. Med. Imag. **14**, 442–453.
8. Kaus, M.R., Warfield, S.K., Nabavi, A., Black, P.M., Jolesz, F.A., Kikinis, R. (2001) Automated segmentation of MR images of brain tumors. Radiology **218**, 586–591.
9. Kruggel, F., Lohmann, G. (1996) BRIAN (brain image analysis) - a tool for the analysis of multimodal brain data sets. In: Proc. CAR 1996, 323–328.
10. Lee, J.H., Garwood, M., Menon, R., Adriany, G., Andersen, P., Truwit, C.L., Ugurbil, K. (1995) High contrast and fast three-dimensional magnetic resonance imaging at high fields. Magn. Reson. Med. **34**, 308–312.
11. Leemput, K.V., Maes, F., Bello, F., Vandermeulen, D., Colchester, A.C.F., Suetens, P. (1999) Automated segmentation of MS lesions from multi-channel MR images. In: Proc. MICCAI 1999, Lecture Notes in Computer Sciences **1679**, 11–21.
12. Loncaric, S., Dhawan, A.P., Broderick, J., Brott, T. (1995) 3D image analysis of intracerebral brain hemorrhage. Comp. Meth. & Prog. Biomed. **46**, 207–216.
13. Loncaric, S., Dhawan, A.P., Cosic, D., Kovacevic, D., Broderick, J., Brott, T. (1999) Quantitative intracerebral brain hemorrhage analysis. In: Proc. SPIE **3661,** 886–894.
14. Maksimovic, R., Stankovic, S. Milovanovic, D. (2000) Computed tomography image analysis: 3D reconstruction and segmentation applying active contour models - snakes. Int. J. Med Info. **59**, 29–37.
15. Moon, N., Bullitt, E., van Leemput, K., Gerig, G. (2002) Model-based brain and tumor segmentation. In: Proc. ICPR 2002, 528-531.

16. Rencher, A.C. (1997) Methods of multivariate analysis. Wiley, New York.
17. Stocker, A., Sipilä, O., Visa, A., Salonen, O., Katila, T. (1996) Stability study of some neural networks applied to tissue characterization of brain magnetic resonance images. In: Proc. ICPR 1996, 472–476.
18. Tao, Y., Grosky, W.I., Zamorano, L., Jiang Z., Gong, J. (1999) Segmentation and representation of lesions in MRI brain images. In: Proc. SPIE **3661**, 930–939.
19. Velthuizen, R.P., Clarke, L.P., Phuphanich, S., Hall, L.O., Bensaid, A.M., Arrington, J.A., Greenberg, H.M., Silbiger, M.L. (1995) Unsupervised measurement of brain tumor volume on MR images. J. Magn. Reson. Imag. **5**, 594–605.
20. Willems, G., Pison, G., Rousseeuw, P.J., van Aelst, S. (2002) A robust Hotelling test. Metrika **55**, 125–138.
21. Zijdenbos, A.P., Dawant, B.M., Margolin, R.A., Palmer, A.C. (1994) Morphometric analysis of white matter lesions in MR images: method and validation. IEEE Trans. Med. Imag. **13,** 716–724.
22. Zijdenbos, A.P., Forghani, R., Evans, A.C. (1998) Automatic quantification of MS lesions in 3D MRI brain data sets: validation of insect. In: Proc. MICCAI 1998, Lecture Notes in Computer Sciences **1496,** 439–448.

Imaging Support of Minimally Invasive Procedures

Terry M. Peters

Robarts Research Institute and University of Western Ontario
London, Ontario, Canada N6A 5K8
tpeters@imaging.robarts.ca

Abstract. Since the discovery of x-rays, medical imaging has played a major role in the guidance of surgical procedures, and the advent of the computer has been a crucial factor in the rapid development of this field. As therapeutic procedures become significantly less invasive, the use of pre-operative and intra-operative images to plan and guide procedures has gained increasing importance. While image-guided surgery techniques have been in use for many years in the planning and execution of neurosurgical procedures, more recently endoscopically-guided approaches have made minimally invasive surgery feasible for other organs. The most challenging of these is the heart. Although some institutions have installed intra-operative, real-time MRI facilities, these are expensive and often impractical. So a major area of research has been the registration of pre-operative images to match the intra-operative state of organs during surgery, often with the added assistance of real time intra-operative modalities such as ultrasound and endoscopy. This paper examines the use of medical images, often integrated with electrophysiological measurements, to assist image-guided surgery in the brain for the treatment of Parkinson's disease, and discusses the development of a virtual environment for the planning and guidance of epi- and endo-cardiac surgeries for coronary artery bypass and atrial fibrillation therapy.

1 Introduction

Minimally invasive surgical procedures are becoming increasingly common, and as a result, the use of images registered to the patient, is a prerequisite for both the planning and guidance of such operations. While many invasive procedures, (traditional coronary artery bypass for example) require relatively minor surgical intervention to effect the desired therapy or repair, the patient is often severely physically traumatized in the process of exposing the site of the therapeutic target. In one sense the objective of minimally invasive approaches is to perform the therapy without the surgery!

Minimally invasive techniques have been in use now for many years, particularly in the brain and skeletal system. The targets in these cases are relatively rigid, making the process of registering pre-operative images to the patient fairly straightforward. For other organs, for example the heart, liver, and kidney, registration is not as simple, and it is these organs that present the major challenges for imaging during minimally invasive surgery.

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 19-26, 2004.

2 Neuro Applications

2.1 Frame-Based Stereotactic Deep-Brain Surgery

Computerized surgical planning systems made their debut in the early 1980's using simple programs that established coordinate systems in the brain based on frame-based fiducial markers. This approach rapidly evolved to allow images from multiple modalities to be combined, so that surgical planning could proceed using information from a combination of MRI, CT, and angiographic and functional images. Such multi-modality imaging was considered important for certain procedures, such as the insertion of probes or electrodes into the brain for recording or ablation, and the ability to simultaneously visualize the trajectory with respect to images of the blood vessels and other sensitive areas. Multi-modality imaging enabled the pathway to be planned with confidence [1;2]. Much of Stereotactic neurosurgery was concerned with procedures involving the safe introduction of probes, cannulae or electrodes into the brain.

2.2 Frameless Stereotaxy

Because the attachment of a frame to the patient's skull is itself invasive, there has been a general desire to eliminate the use of the frame from the procedure. However, without the frame to provide the fiducial markers, some other type of reference system must be employed to register the patient to the image(s). A commonly used registration method is point-matching, where homologous landmarks are identified both in the images and on the patient. Unfortunately, some variation in the identified locations of the landmark points on the patient is always present, and it is difficult to pinpoint exactly the same locations within the patient's three-dimensional image. Point matching is often employed in conjunction with surface-matching, which is achieved using the probe to sample points on the surface of the patient, and then determining the best match of this point-cloud to an extracted surface from the 3-D patient image. Under ideal conditions, accuracy approaching that available with Stereotactic frames can be achieved [3].

2.3 Integration of Physiological Information with Images

A common treatment for Parkinson's disease involves the ablation or electrical stimulation of targets in the deep brain, either within the thalamus, the sub-thalamus, or the globus pallidus. The standard imaging modality for guiding the surgical treatment of targets in the deep brain is a T1-weighted volumetric MR image. This image however does not show the affected parts of the brain directly, nor does it demonstrate the deep-brain nuclei that constitute the targets for such therapy. Other means must be used to define the target areas within the otherwise homogeneous regions. Approaches to solve this problem include the use of atlases mapped to the patient images, using linear, piece-wise linear, or non-rigid registration. This is often complemented with information gained from electrophysiological exploration of the

target area, together with anatomical data provided by MRI, CT, or angiography. Mapping the electrophysiological responses recorded during such procedures onto the 3-D anatomical images of the patient helps the surgeon navigate towards the desired target. Moreover, a database of such responses, normalized through non-rigid image registration to a standard data space, can be integrated with the patient's 3-D MRI to assist the surgeon by predicting the likely target for creating a therapeutic lesion [4]. A typical example of electrophysiology integrated with 3D MRI for guiding the surgeon to the target in the treatment of Parkinson's disease is shown in Figure 1.

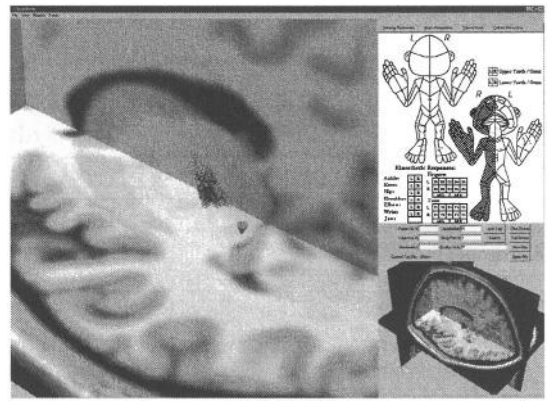


Fig. 1. Example of electrophysiological database acquired from multiple patients, and integrated with 3D MRI of patient. The figure in the upper right is the interface whereby the user selects the body-region associated with the stimulus/response data being entered or displayed

2.4 Brain-Shift Compensation

If the entry point for the target within the brain is inserted into the otherwise intact skull, the brain may be treated as a rigid body and pre-operative images, registered to the patient, are appropriate for guidance during the procedure. In the presence of a craniotomy however, significant brain shift occurs, and the pre-operative images no longer represent the intra-operative morphology of the brain. Various approaches have been used to solve this problem, from the use of MR imaging systems that are somewhat "operating-room unfriendly", to intra-operative ultrasound integrated with the image-guidance protocol. Updating of neuro MR volumes using intra operative ultrasound continues to be an active research topic in our laboratory and others.

As the reach of minimally-invasive surgery extends beyond the brain, so the demands on image processing to accommodate procedures in other organ systems increases. Most of these procedures involve non-static states, being affected by blood-pressure changes, breathing or the interaction with surgical tools. If image-guidance is to be used in these situations, realistic models that are synchronized in space and time with the actual patient organ are required.

3 Application to the Heart

In being an appropriate candidate for image-guided surgery, the heart is probably at the opposite end of the spectrum from the brain. Despite this, we were motivated to attempt the goal of developing a dynamic cardiac model for planning and guidance purposes by our surgical colleagues who are performing robotically-assisted coronary bypass surgery, as well as electro-ablative procedures within the left atrium.

3.1 Bypass Surgery

Many conventional cardiac surgical procedures require a sternotomy and a cardiopulmonary bypass procedure, which subjects patients to significant trauma and lengthy hospital stays. Recently, minimally invasive direct coronary artery bypass (MIDCAB) procedures have been introduced, performed via instruments introduced into the chest via trochars and guided endoscopically. Such techniques are making a significant impact on cardiac surgery, but because of the extreme difficulty in manipulating the instruments at the distal ends of the trochars, several surgical teams have begun to perform coronary bypass surgery on the beating heart in the intact chest, using tele-operated robots inserted into the chest via intercostal ports [5].

In spite of the sophistication of these robotically assisted systems, the use of medical images in the planning of the procedure is mostly limited to conventional chest-radiographs and angiograms. The use of such simple images makes it extremely difficult to plan the positions for the entry ports between the ribs, and provides minimal guidance during the procedure.

While minimally invasive and robotically assisted approaches are enjoying increasing application, the potential benefits have not yet been fully realized. In the absence of a more global perspective of the target organ and its surroundings, the spatial context of the endoscopic view can be difficult to establish. Other intrinsic limitations of the endoscope include its inability to "see" beneath the surface of the target, which is often completely obscured by bleeding at the surgical site. To assist the planning and guidance of such procedures, there are a number of reports [6;7;8] describing the development of static virtual modeling systems to plan cardiac surgical procedures.

3.2 Atrial Fibrillation Surgery (AFS)

Arrhythmias have long been controlled with minimally invasive approaches, in both the operating room and in the electrophysiology (EP) laboratory using catheter techniques.

Atrial fibrillation is difficult to treat using catheter techniques, but a conventional surgical approach is considered excessively invasive. Colleagues at the London Health Sciences Centre, London, Canada, have recently developed a minimally invasive technique that permits ablative therapies to be performed within the closed beating heart, using instrumentation introduced through the heart wall. This duplicates the surgical procedure that is otherwise performed using an open heart technique, with

life-support being provided by a heart-lung machine. However, this approach requires the support of image guidance, to integrate both anatomical and electrophysiological data, highly analogous to the neuro-physiological mapping approaches described earlier.

Unlike during epicardial procedures, an endoscope cannot be used to navigate inside the blood-filled atrium. A fully minimally invasive approach requires the use of surrogate images based on a simulated heart model dynamically mapped to the patient. This model must in turn be complemented with the intra-procedure EP data mapped onto the endocardial surface and integrated within the virtual planning and guidance environment.

4 Towards a Virtual Cardiac Model

Effective image guidance for these surgical procedures is challenging and demands a cardiac model that is registered to the patient in both space and time. The image-guided surgery laboratory at the Robarts Research Institute in London, Canada is currently engaged in a project to achieve this goal via the following steps:

1. The creation of a dynamic cardiac model
2. Registration of the model to the patient
3. Synchronization of the model to patient
4. Integration of patient and virtual model
5. Integration of registered intra-cardiac and intra-thoracic ultrasound
6. Tracking of tools and modeling them within virtual environment.

4.1 Dynamic Cardiac Model

Dynamic imaging using both MR and CT has become a common feature of contemporary medical imaging technology, but it is difficult to acquire high quality images at every phase of the cardiac cycle. However, during end-diastole, one can obtain a quasi-static 3D image of relatively high quality. While images acquired at other phases of the cardiac cycle are noisier and often contain motion artifacts, they nevertheless contain much of the information necessary to describe the motion of the cardiac chambers throughout the heart cycle. Capturing this information and applying it to the high-quality static image allows an acceptable dynamic model to be created. This behavior has been exploited by Wierzbicki *et al.* [9] to generate high quality dynamic image models from patient data that can be incorporated in a dynamic virtual model of the heart within the thorax.

Within such a virtual environment, it is also important to integrate data from tracked real-time imaging tools, such as endoscopes and ultrasound probes. Our work in this area has recently been reported by Szpala *et al.* [10] who demonstrated that the dynamic dataset representing the virtual cardiac model could be integrated with the real-time endoscopic image delivered by a tracked endoscope.

Our work continues to refine these techniques, as well as to address the problems of image-to-patient registration; tracking intra cardiac and intra thoracic ultrasound, the mapping of cardiac electro-physiology into the model environment, and the representtation of tracked tools within the virtual environment.

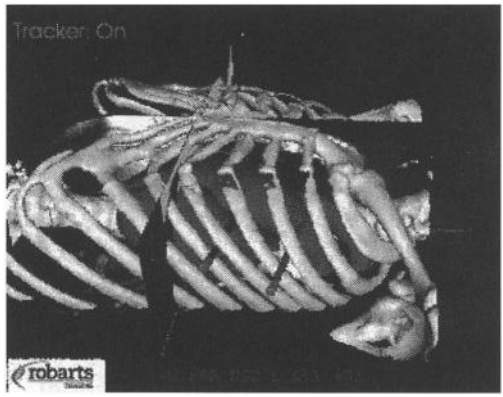


Fig. 2. Virtual model of beating heart within thorax with integrated representation of robotic probes.

5 Challenges

There are many challenges associated with this endeavour, and they are not unique to the application for cardiac therapy. The most pressing is perhaps the development of means to rapidly deform the dynamic organ models in response to the intervention of a therapeutic instrument. This entails not only endowing the model with sufficiently realistic characteristics to allow it to behave appropriately, but also to ensure that performance is not compromised in the process. Finite element representations of organs have been proposed by many groups, and well characterized models can predict tissue behaviour accurately. However, it is acknowledged that finite element model (FEM) techniques are often orders of magnitude too slow for real-time operation, and that alternative approaches must be developed. One method is to parameterize the behaviour of tissues based upon observed responses of actual or finite-element models of organs to sample stimuli [11; 12]. Another challenge will be to enable the updating of the model environment rapidly as intra-operative imaging detects the changes during the procedure. This will require accurate tracking of the intra-operative imaging modality, rapid feature mapping between the image and the model, and local deformation of the model to match the image. While these operations require a large computational overhead of multiple simultaneous execution modules, we believe that the evolving levels of readily-available computational power will be sufficient to accomplish these goals in the near future.

On a broader front, a working group discussing the future of intraoperative imaging at a recent workshop[1] held in Maryland, USA April 18-20 2004, identified a number of challenges that must be met before the ideas presented here, and the ubiquitous use of image-guided intervention in general, can become established on a routine basis. It was observed that most operating rooms in the world are not even equipped with PACS, let alone the infrastructure to bring sophisticated 3D and 4D imaging into the OR suite; that we still lack the tools to rapidly pre-process images (i.e. segment,

[1] OR 2020 http://www.or2020.org/

mesh) in an automatic pipeline fashion; that metrics for success of both the technology and outcomes related to new image-guided surgical procedures are poorly defined at present, and that there remains a great deal of incompatibility across manufactures with respect to standard interfaces to equipment.

The workshop presented a number of "Grand Challenges" that the members considered on which industry should focus to enable these technologies:

1. the development of integrated displays that can inherently handle multiple modalities in multiple formats simultaneously;
2. that systems be designed to accommodate inherent tracking and registration across modalities, tools, endoscopes, microscopes;
3. that advanced non-rigid image registration, at both the pre-op and intra-operative phases of the procedure be developed together with appropriate error measures; and
4. that OR-destined imaging systems should be developed from the ground up, rather than as diagnostic systems retrofitted in the OR.

I believe that these issues MUST be addressed in a coordinated fashion, with full participation of industry, if we as a community are to make significant progress in the development of image-guidance to enhance minimally invasive procedures.

Acknowledgement. This work was supported by grants from the Canadian Institutes of Health Research (CIHR MOP 14735, MOP 62716); the Heart and Stroke Foundation of Ontario (NA 4755), and the Canadian Foundation for Innovation. The contributions of the following individuals are gratefully acknowledged: Dr D Gobbi, Dr Y Starreveld, Dr K Finnis, Dr S Szpala, Dr. G Guiraudon, Dr M Drangova, Mr M Wierzbicki, Dr M Wachowiak, Mr J Moore, Ms G-A Turgeon, Mr Nick Hill, Mr Qi Zhang, Mr X Huang, Dr H Zhong.

References

1. Henri, C. J., Collins, D. L., and Peters, T. M., Multi-modality image integration for stereotactic surgery planning, *Medical Physics,* no. 18, pp. 166-176,1990.
2. Peters T.M., Henri C.J., Munger P., Takahashi A.M., Evans A.C., Davey B., and Oliver A., Integration of stereoscopic DSA and 3D MRI for image-guided neurosurgery, *Computerized Medical Imaging and Graphics,* vol. 18, no. 4, pp. 289-299, 1994.
3. Zinreich, S. J., Tebo, S. A., Long, D. M., Brem, H., Mattox, D. E., Loury, M. E., vander, Kolk, C. A., Koch, W. M., Kennedy, D. W., and Bryan, R. N., Frameless stereotaxic integration of CT imaging data: accuracy and initial applications, *Radiology,* vol. 188, no. 3, pp. 735-742, Sept.1993.
4. Finnis K.W., Starreveld Y.P., Parrent A.G., Sadikot A.F., and Peters T.M., A three-dimensional atlas of subcortical electrophysiology for the planning and guidance of functional neurosurgery, *IEEE Transactions on Medical Imaging,* vol. 21, no. 1, pp. 93-104, 2003.

5. Damiano, R. J., Jr., Ehrman, W. J., Ducko, C. T., Tabaie, H. A., Stephenson, E. R., Jr., Kingsley, C. P., and Chambers, C. E., Initial United States clinical trial of robotically assisted endoscopic coronary artery bypass grafting, *Journal of Thoracic & Cardiovascular Surgery,* vol. 119, no. 1, pp. 77-82, Jan.2000.
6. Adhami, L. and Coste-Maniere, E. A versatile system for computer integrated mini-invasive robotic surgery. Dohi T and Kikinis R. MICCAI 2002 LNCS vol 2488, 272-281. 2002. Heidelberg, Springer..
7. Lehmann G.L., Chiu A, Gobbi D., Starreveld Y.P, Boyd D.W., Drangova M., and Peters T.M. Towards dynamic planning and guidance of minimally invasive robotic cardiac bypass surgical procedures. Niessen, W. J. and Viergever, M. Proc MICCAI 2001, LNCS vol. 2208, 368-375. 2001.
8. Chiu A.M., Dey D., Drangova M., Boyd WD, and Peters T.M., "3-D Image Guidance for Minimally Invasive Robotic Coronary Artery Bypass (MIRCAB), *Heart Surgery Forum,* vol. 3, no. 3, http://www.hsforum.com/vol3/issue3 /2000-9732.html, 2000.
9. Wierzbicki M., Drangova M., Guiraudon G., and Peters T.M., Validation of dynamic heart models obtained using non-linear registration for virtual reality training, planning, and guidance of minimally invasive cardiac surgeries, *Medical Image Analysis,* vol. 8, no. 3, pp. In Press, 2004.
10. Szpala S., Guiraudon G., and Peters T.M. Cardiac endoscopy enhanced by dynamic organ modeling for minimally-invasive surgery guidance. Ellis R.E. and Peters T.M. Proc MICCAI 2003, LNCS vol 2878, 499-506. 2003.
11. Delingette, H., Cotin, S., and Ayache, N., Efficient linear elastic models of soft tissues for real-time surgery simulation, *Studies in Health Technology & Informatics,* vol. 62 pp. 100-101, 1999.
12. Montagnat, J., Delingette, H., and Ayache, N., A review of deformable surfaces; topology, geometry, deformation., *Image and Vision Computing,* vol. 19, no. 14, pp. 1023-1040, 2001.

Hands-On Robotic Surgery: Is This the Future?

Brian L. Davies, Simon J. Harris, Ferdinando Rodriguez y Baena, Paula Gomes, and Matjaz Jakopec

Mechatronics in Medicine, Mechanical Engineering Department, Imperial College London, Exhibition Road, London SW7 2AZ, UK
b.davies@imperial.ac.uk

Abstract. An introduction to robotic surgery is given, together with a classification of the range of systems available with their problems and benefits. The potential for a new class of robot system, called a hands-on robot is then discussed. The hands-on robotic system, which is called Acrobot®, is then presented for total knee replacement (TKR) surgery and for uni-condylar knee replacement (UKR) surgery. CT-based software is used to accurately plan the procedure pre-operatively. Intra-operatively, the surgeon guides a small, special-purpose robot, which is mounted on a gross positioning device. The Acrobot® uses active constraint control, which constrains the motion to a predefined region, and thus allows the surgeon to safely cut the knee bones to fit a TKR or a UKR prosthesis with high precision. A non-invasive anatomical registration method is used. The system has undergone early clinical trials of a TKR surgery and, more recently a blind randomised clinical trial of UKR surgery. Preliminary results of the UKR study are presented in which the pre-operative CT based plan is contrasted with a post operative CT scan of the result, in an attempt to gain an objective assessment of the efficacy of the procedure. Finally, proposals for future requirements of robotic surgery systems are given.
Keywords: Robotic surgery; Medical robotics; Active constraint control.

1 Introduction

The use of medical robots is a relatively recent phenomena. It started in the mid 1980s with the use of industrial robots which were used as a fixture to hold tools at an appropriate location and orientation for neuro-surgery. In this application, having arrived at the appropriate location, power was removed from the robot, and the surgeon manually carried out simple tasks such as drilling the skull. Thus the robot acted as a purely passive positioning system. Subsequently industrial robots were applied for orthopaedic surgery and modified for safe use. These robots were used in an autonomous mode, so that they carried out a pre-planned sequence of motions with the surgeon acting as an observer, who would only intervene in an emergency to press the off-button. This autonomous mode worked best for orthopaedic surgery because the leg could be clamped as a fixed object, so that the robot acted in the same way as a standard" computer numerical control" machining process. Clearly if the tissue moved during the

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 27–37, 2004.

procedure, then a pre-planned autonomous system, no matter how accurate, would be useless since the location of the target was unknown. Thus for all soft tissue surgery, where tissue can deform or collapse during the cutting process, autonomous systems are less appropriate.

An example of a special-purpose autonomous system developed by the author is the Probot, for prostate resection surgery. (Fig 1(a)). This involved a framework to resect a series of overlapping cones from the prostate, to remove a blockage from the urethral duct caused by a benign adenoma. The special purpose robot was placed on a gross positioning system so that when the cutter was in the appropriate location, the positioning system was locked off for safety, thus enabling a small low-force robot to be used for the resection process [1]. A hot wire diathermic loop was used, together with a conventional resectoscope, to chip away segments of prostate tissue. This series of repetitive cuts is very tiring for an urologist in conventional surgery and results in disorientation, requiring frequent re-referencing by the surgeon to know where the tool is located within the prostate. At the start of the procedure, to image the gland, the resectoscope was replaced by a standard transurethral ultrasound probe. The robot was scanned along the axis of the prostate, generating a series of 2D views that were captured by a computer to form a 3D model. This model was then used as a database for instructing the range of motions of the robot, to ensure the appropriate resection. Subsequently replacing the ultrasound probe with the resectoscope ensured that the cutting and imaging processes could take place using the same referencing system, thus ensuring accuracy to within a few millimetres. Fortunately the prostatectomy process is one primarily of debulking to remove the urological obstruction, and so this accuracy figure for the autonomous robot was acceptable.

One difficulty with these autonomous robots is that it is unclear who is in charge of the procedure. Is it the surgeon or is it the robot programmer? The programmer may have incorporated a number of conditional motions of which the surgeon has no knowledge. This concern, and suspicion that they are not in charge, has caused a number of surgeons to refrain from robotic surgery and instead they have preferred to adopt computer aided surgery in the form of "navigation tracking systems. It was to accommodate such concerns that the author's group has developed the concept of "hands-on" robots for surgery, which will be described more fully later.

Another alternative to hands-on robots are telemanipulator systems. These involve a master controller operated by the surgeon, who is generally alongside the slave robot acting on the patient in the operating room. Usually such telemanipulators utilise very good vision, but have little haptic sense. An example of this is the Da Vinci system by Intuitive Surgical for closed heart surgery, which uses an excellent 3D endoscopic camera for three dimensional display at the master [2]. The Da Vinci system has two separate slave systems for tools, which act through small ports in the chest and utilise scissors and grippers that can rotate and pivot to manipulate as if a wrist were placed inside the body. This dexterity gives extremely good manipulative capability inside the body, without the need

to have massive openings in the chest wall. These telemanipulator systems generally operate as open loop controlled structures, in which the surgeon visually checks the location of the tool and target tissue and then closes the control loop to ensure that the tool acquires the target. Thus it is the surgeon who can adapt to distorting and moving tissue, in cases when an autonomous system would be inadequate. The difficulty with a telemanipulator is that it requires considerable concentration from the surgeon, and even though magnification and scaling of motions is possible, the procedure is very tiring and time-consuming when only vision is available with no sense of touch.

A hands-on robot was designed by the author as a special-purpose system intended for orthopaedic surgery in which a gross positioning system allows a small active robot to be placed in the region of interest. The gross positioner is then locked off for safety, leaving only a small low-force robot to interact directly with the patient. A force controlled input handle is located near the tip of the robot and can be held by the surgeon to move the robot around under force control. A high-speed rotary cutter is placed at the tip of the robot so that it can machine bones very accurately. The robot can thus be pre-programmed to move within a safe constrained region so that damage to critical structures, such as ligaments, can be avoided. This ability to actively constrain cutting motions has given rise to the term Active Constraint Robot and the surgical system is called the ACROBOT, which has given rise to the University spin-off company, the Acrobot Company Ltd.

2 Total Knee Replacement (TKR) Surgery and Uni-condylar Knee Replacement (UKR)

Total knee replacement (TKR) surgery and uni-condylar knee replacement (UKR) surgery are common orthopaedic procedures to replace damaged surfaces of the knee bones with prosthetic implants. Typically, TKR and UKR prostheses consist of three components, one for each knee bone: tibia, femur and patella (see Fig 1(b)). To fit the prosthesis, each of the knee bones is cut to a specific shape (usually a set of flat planes) which mates with the mounting side of the corresponding prosthesis component. To ensure normal functionality of the knee, and long-lasting, pain-free implant, all components must be placed onto the bones with high precision, both with regard to the bone axes, and with regard to the mating surfaces between the bone and prosthesis. Conventionally, the surgeon cuts the bones using an oscillating saw, together with a complex set of jigs and fixtures, in an attempt to accurately prepare the surfaces of the bones. It is very difficult to achieve a good fit and proper placement of the prosthesis, even for a skilled surgeon using state-of-the-art cutting tools and fixtures. A conventional TKR study [3] reports a deviation from ideal prosthesis alignment greater than 9° in 7% of cases, and greater than 5° in 34% of cases. The reason for this is that the cutting tools and fixtures are made for an average human anatomy, and their placement largely depends on surgeon's experience. Furthermore, the fixtures are used sequentially, which can result in an accumu-

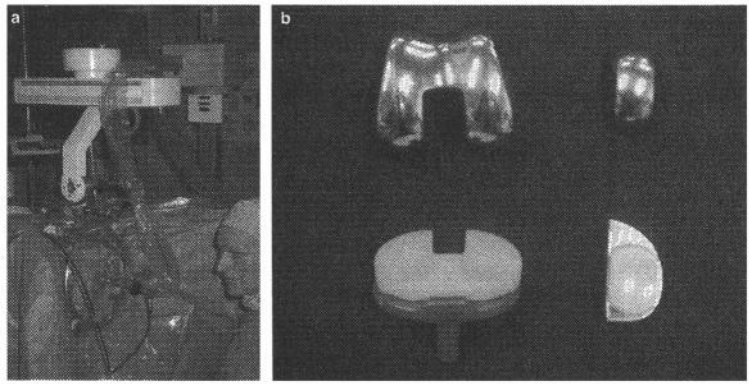


Fig. 1. (a) Probot**(b)** TKR prosthesis (left) and UKR prosthesis (right)

lation of errors. Another source of error lies in the oscillating saw, as its blade tends to bounce off a hard part of the bone despite the guiding tools, which can result in a poor surface finish. UKR procedures are even more demanding than TKR, due to the difficulty of access through small incisions. To overcome the problems of the conventional TKR/UKR surgery and improve the results of the procedure, a robotic surgery system is being developed at Imperial College, and has undergone early clinical trials [4,5]. A prosthesis alignment of less than 2° and a sub-millimetre placement accuracy have been achieved with the robotic assisted approach. The system consists of a pre-operative planning workstation and an intra-operative robotic system.

3 Pre-operative Planning

The planning stage of the Acrobot system is much more thorough than the planning for a conventional TKR/UKR surgery (which simply involves placing prosthesis templates over x-ray images). First, a CT (computed tomography) scan of the patient's leg is taken. The interactive planning software, developed at Imperial College, is used to process CT images and plan the procedure: 3D models of the knee bones are built and the bone axes are determined. The surgeon then interactively decides the prosthesis type, size and placement (Fig 2). The software provides a number of different 3D views of the leg and the implant, to help the surgeon plan the procedure. Once the prosthesis model is in the correct position over the bone model, the planning software generates the constraint boundaries, which are then transferred to the intra-operative robotic system.

4 ACROBOT® Robotic System

In contrast to other robotic systems for orthopaedic surgery, such as Robodoc [6] or Caspar [7], which use modified industrial robots, a small, low-powered, special-purpose robot, called Acrobot® has been built for safe use in a crowded sterile

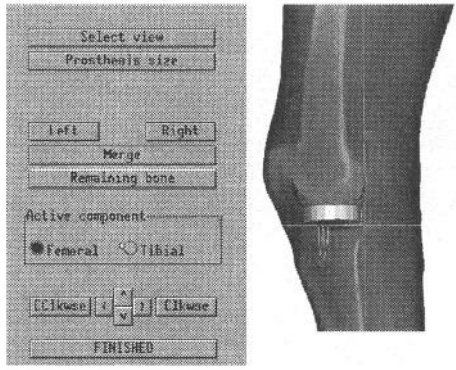


Fig. 2. User interface of the pre-operative planning software

operating theatre environment. The Acrobot is a spherical manipulator with three orthogonal axes of motion: Yaw, Pitch and Extension (Fig 3). It has a

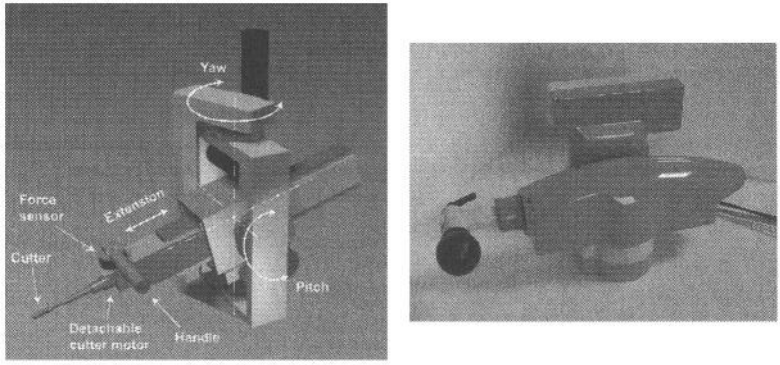


Fig. 3. Acrobot Head and its latest implementation

relatively small reach (3050cm) and range of angles (-30°+30°), which ensures accurate operation with low-powered motors. Consequently, the robot is relatively safe, because potential damage is limited in terms of force and is constrained to a small region. The mechanical impedance of the axes is low and similar for all axes, allowing the robot to be moved by the surgeon with low force. The surgeon moves the robot by pushing the handle near the tip of the robot. The handle incorporates a 6-axes force sensor, which measures the guiding forces and torques. This force/torque information is used in active constraint control of the robot. A high-speed orthopaedic cutter motor is mounted at the tip of the Acrobot. Different rotary cutters and tools can be repeatably mounted into the motor. For sterility reasons, the cutter motor is placed in a special mount, which allows the motor to be removed and autoclaved before the surgery, and placed

onto the robot at the start of the surgery. Due to its small working envelope, the Acrobot is placed on a 6-axes gross positioning device (Fig 4), which moves

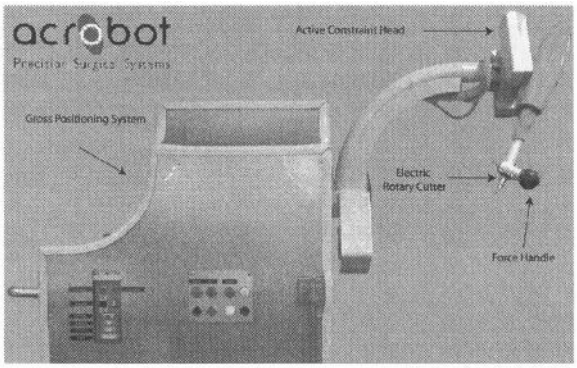


Fig. 4. The Acrobot mounted on a gross positioning device

the Acrobot to different optimal cutting locations around the knee. The velocities of the axes are limited, and secondary encoders are fitted on each axis as a check for increased safety. Furthermore, the device is locked off when the bone is machined, and is only powered on for a short period between cutting two consecutive planes to move the Acrobot to the next optimal cutting position. The gross positioning robot is mounted on a wheeled trolley, to allow the system to be quickly placed near the operating table or removed from it. The trolley wheels can be locked. In addition, two clamps are provided on the trolley, that clamp onto the side rails of the operating table, and thus rigidly link the robot base to the operating table. The Acrobot and the gross positioning device are covered with sterile drapes during the surgery, with the sterile cutter motor protruding through the drapes. Sterilisable leg fixtures are another important part of the robotic system, as they ensure that the bones do not move with respect to the robot base during the procedure. Two special bone clamps are rigidly clamped to the exposed parts of the tibia and femur. Each of the bone clamps is linked to a base frame (attached to the side rails of the table) with three telescopic arms (Fig 5(a)). The ankle is placed into a special foot support mounted on the operating table, whereas the weight of the patient has proven to be enough to immobilise the femur at the hip joint.

5 Active Constraint Control

The Acrobot uses a novel type of robot control - active constraint control [8,9]. The surgeon guides the robot by pushing on the force controlled handle at the tip of the robot, and thus uses his/her superior human senses and understanding of the overall situation to perform the surgery. The robot provides geometric accuracy and also increases safety by means of a predefined 3D motion constraint.

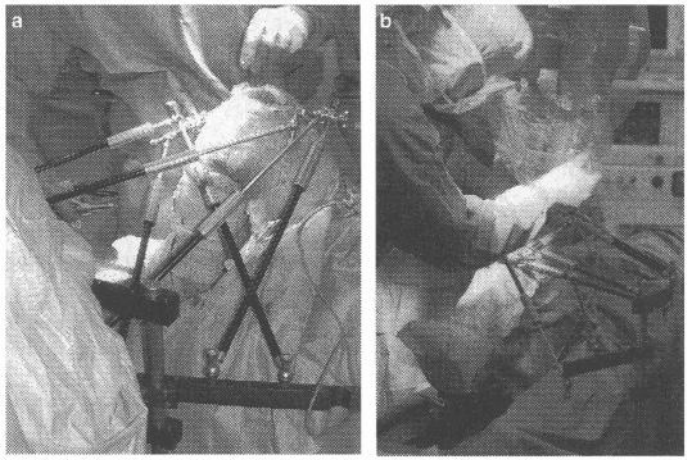


Fig. 5. (a) Leg fixtures in place during a clinical trial **(b)** Clinical application of UKR

This "hands-on" approach is very different from other surgical robots and is favoured by the surgeons, as the surgeon is kept in the control loop. The basic idea behind active constraint control is to gradually increase the stiffness of the robot as it approaches the pre-defined boundary. In other words, the robot is free to move inside the safe region (RI), due to low mechanical impedance of the robot. As a result, the surgeon can feel cutting forces at the tip of the robot. At the boundary of the safe region, the robot becomes very stiff, thus preventing further motion over the boundary (RIII). In addition, the portion of the guiding force normal to the boundary is directly compensated, which substantially reduces the over-boundary error (this error is inevitable due to the limited motor power). To avoid instabilities at the boundary, the stiffness increases gradually over a small region close to the boundary (RII). Furthermore, only the stiffness in the direction towards the boundary is adjusted, to allow motion along or away from the boundary with a very low guiding force. This is achieved by a two level control. The inner loop (2 ms) is a joint position/velocity control loop with friction, gravity and guiding force compensation. The outer, slower loop (approx. 10-15 ms, depending on complexity of the boundary) is the boundary controller, which is executed in Cartesian space and adjusts the parameters of the inner loop, according to the robot's position and the surgeon's guiding force. Because of the nature of the prosthesis used, which requires a number of flat planes to be cut, the constraint boundary is defined as a "2.5D" boundary, formed from a plane and a closed 2D outline. The plane part allows a flat plane to be cut into the bone, whereas the 2D outline part provides protection for the surrounding tissue. Each of the two parts can be controlled separately. Considering the plane part, the stiffness is adjusted according to the distance to the plane. For the 2D outline, the nearest point on the outline is found (when projected onto the plane). The stiffness of the robot is then adjusted independently in normal and

tangential directions to the outline at this nearest point. These stiffnesses, each of which is computed along three orthogonal axes, are then combined before being passed to the inner loop.

6 Registration

Registration is a process to determine the geometrical relationship between different types of information about the patient and the surgical tools. In other words, the registration transformation R has to be determined from the coordinate system of the CT scanner (where pre-operative planning data are defined) to the base coordinate system of the robot (relative to which the locations of the bones and the robot's tip are defined). The planning data are then transformed into the coordinate system of the robot. At an early stage of the project, when trials were performed on plastic phantoms and cadavers [10], a registration method using fiducial markers was used. This involved placing 4 marker screws prior to taking a CT scan of the leg. The markers represent 4 corresponding non-colinear points in pre-operative (the marker centre is detected in the CT images) and robot's coordinate systems (the marker is touched by the tip of the robot intra-operatively), from which the transformation R can be computed. As this method requires an additional surgery to implant the markers, it was regarded as unacceptable for clinical trials. A non-invasive anatomical registration method was implemented. The method is based on the iterative closest point (ICP) algorithm [11,12], which determines the registration transformation by iteratively matching a set of surface points in one coordinate system to a corresponding surface in another coordinate system. The procedure is as follows: first, the surgeon acquires four landmark points, with a special 1 mm diameter ball probe mounted into the cutter motor. These four points are used to compute the initial registration estimate for the ICP algorithm. The surgeon then acquires a set of randomly selected points (typically 20–30) on the exposed bone surface. The ICP algorithm then registers the bone by matching this set of points to the pre-operative bone surface model.

7 Clinical Application

After being successfully tested on plastic phantom bones and cadavers, the Acrobot UKR system was brought into the operating theatre for a blind randomised clinical trial under MHRA committee approvals clinical trials (Fig 5(b)). 13 conventional and 13 robotic clinical trials have been performed. The reason for undertaking a UKR trial, rather than a TKR, was that there is more interest in the Minimally Invasive Surgery (MIS) procedure, which is also regarded as more challenging than TKR. The procedure was as follows: a CT scan of the patient's leg was taken a few days before the surgery. The procedure was planned using the planning software and the data for the robotic system were generated. Intraoperatively, the knee was exposed and the bones were immobilised using the leg fixtures. In the mean time, the robot was set up and covered with sterile drapes.

Once the leg was immobilised, the robot was wheeled next to the operating table and clamped to it. The femur and the tibia were then registered (with a 1 mm diameter ball probe) and cut (with a 0.8 mm diameter ball cutter). The patella was prepared manually, as the conventional procedure is regarded to be accurate enough, and the potential accuracy improvement does not justify the added complexity of the robot-assisted procedure. The robotic system and leg fixtures were then removed from the operating table, and the procedure was finished as in conventional TKR surgery: the knee motion was tested with a trial prosthesis, the prosthesis components were fitted onto the bones, and the incision was closed. In all cases, the surgeon was able to succesfully register the bones, which was indicated by a low RMS error (less than 0.4 mm) of the registration points, and confirmed by real-time display. In all cases involving cutting the bone with the aid of the robot, the fit of the prosthesis components onto the bone was very good. Furthermore, the components were found to mate correctly, giving proper bone alignment and a good range of motion.

In addition to the TKR procedure, the Acrobot system has now completed a blind randomised clinical trial, in which 13 patients underwent conventional uni-condylar knee surgery and 13 a robotic uni-condylar surgery. In order to be consistent, both groups were subject to a preliminary CT scan and a computer-based planning procedure in which an appropriate size of prosthesis was chosen. The components were positioned to ensure the correct load lines between hip centre and femoral knee component and the ankle centre and tibial component. Thus in both conventional and robotic cases, a planned procedure was available. Post operatively, both groups were subject to a CT scan and a blind evaluation to check how accurately the plan had been achieved. This accurate measurement process was necessary because it is possible to make claims for accuracy which can be very subjective. By choosing different viewing angles for a radiographic X-ray, when accompanied by knee flexion, it is possible to show good alignment of the prostheses, even though very large errors are actually present. The use of a CT scan to demonstrate accuracy also avoids the need for long-term knee scores, which besides being a "soft" measure of achievement can take many years to prove the benefit of the procedure. Such long-term outcomes do not lend themselves to the shorter term financial survival of the robot supplier company! The blind, randomised, results of the clinical trial have shown that the majority of the robotic cases are within 2° of the plan with only one at 4° and one at 3°. This contrasts with the conventional cases where only half were less than 2° and a number extended to 7°.

The ability to measure each stage of the machining process using the robot has demonstrated that for uni-condylar prostheses, where the prosthesis "footprint" is only 2 cms wide, the use of a cement mantle, even though applied with great care, can on its own give rise to a malalignment of 2°. Thus, since a robot can produce an exact shape to promote bone growth, the use of cementless prostheses with a hydroxy-appetite coating to promote bone growth, avoids the inaccuracies inherent in the cement mantle.

8 Conclusions

Whilst autonomous robot systems have had some success in orthopaedics, their ability to adapt to changing position and shape of the target tissue has been limited. Although telmanipulator systems can adapt to changing position of tissue, the need for the surgeon to close the control loop visually leads to a slow and tiring procedure in which there is no constraining ability for the robot to help the surgeon avoid damaging critical tissue or organs. The use of a hands-on approach, however, not only ensures that the surgeon is constrained to a safe region, but can also provide a degree of accuracy which is not possible conventionally, and would be difficult with computer aided navigation systems. It will, however, be necessary to ensure that the cost and complexity of the resulting integrated robotic system is as low as possible. This is because it is seldom that the robot is performing in a life-critical environment in which there is no alternative. It is more usual that there are benefits which can accrue from the use of the robot, but they must be justified against the inevitable increase in cost and complexity of the total robotic implementation. If this choice is made with care and a knowledge of the best method of implementation, then the use of hands-on robots will undoubtedly have a very bright future. In order to ensure maximum use in future, surgical robots need to be small, low cost, general purpose robots that are also rigid and accurate with a wide reach. They also need to be easy to use in the OR, ie, have a small footprint, be simple to integrate, and not be affected by (or electrically affecting) other OR systems.

They must also allow a high throughput in the OR. Apart from a small number of "life or death" applications, benefits of robotic procedures will only be justifiable, if OR times and staff levels are similar to or less than conventional. There is thus a need to be able to mount a robot system quickly and easily after the patient is on the operating table. Robots must also be simple to use, not requiring additional technical support teams in the OR, since dedicated technical teams in the OR cannot be justified on the grounds of cost and the need for such a team to be continuously available. Since the equipment is more complex than CAS navigation systems it will require a specialist training period beforehand for the surgeon. The robot must have a foolproof set-up procedure, with a transparent and easy to use surgeon/computer interface and a fail-safe system with comprehensive diagnostics.

With regard to safety, researchers and supplying organisations are unsure how safe is "safe", i.e., what is the minimum cost and complexity that are needed for the task (since 100% safety is not feasible, no matter how costly the system). This uncertainty is slowing down the application of systems. National Health and safety organisations are often over-zealous in the absence of agreed guidelines, to avoid being criticized for being too lax. There is thus a need for an international consortia with national representatives to be set up and funded. There is a need for guidelines in the first instance since everyone is wary of yet more "standards", which would be very lengthy and costly to produce.

Provided that these suggestions for future improvements are implemented, there is every reason to expect that hands-on robots will have a long and successfull future.

Acknowledgements

The support of the Acrobot Company Limited is gratefully acknowledged.

References

1. Davies, B. L., Harris, S.J., Hibberd, R. D.: "The Probot- an active robot for prostate resection". Proc. I Mech E, Part H, Vol 211, (H4), (1997) pp 317–326
2. Casula R., Athanasiou T., Darzi A.: "Minimal acess coronary revascularisation without cardiopulmonary bypass - The impact of robotic technology in current clinical practice". Int J Medical Robotics and Computer Aided Surgery, Vol 1, No1, (2004) pp 98–106
3. Tew, M. and Waugh W.: "Tibiofemoral Alignment and the Results of Knee Replacement", Journal of Bone and Joint Surgery, Vol 67-B, No 4, (1985) pp 551–556
4. Jakopec, M., Harris, S. J., Rodriguez y Baena, F., Gomes, P., Cobb, J. and Davies, B. L. (2001) "The First Clinical Application of a "Hands-on" Robotic Knee Surgery System", Computer Aided Surgery, Vol 6, No 6, pp 329–339
5. Cobb, J., Henckel, J., Richards, R., Harris, S., Jakopec, M., Rodriguez, F., Gomes, P.: "Robot assisted minimally invasive uni-compartmental arthroplasty: results of first clinical trials". Proc. CAOS Int. Conf. 04, Chicago, (2004) pp 4–6, ISBN 0-9754971-0-3
6. Mittelstadt, B. D., Kazanzides, P., Zuhars, J. F., Williamson, B., Cain, P., Smith, F. and Bargar, W. L.: "The Evolution of a Surgical Robot from Prototype to Human Clinical Use", in Taylor, R. H., Lavallee, S., Burdea, G. C. and Mösges, R. (Ed.), Computer-Integrated Surgery: Technology and Clinical Applications, The MIT Press, (1996) pp 397–407
7. Siebert, W., Mai, S., Kober, R. and Heeckt, P. F.: "Technique and First Clinical Results of Robot-Assisted Total Knee Replacement", The Knee, Vol 9, No 3, (2002) pp 173–180
8. Ho, S. C., Hibberd, R.D. and Davies, B. L.: "Robot Assisted Knee Surgery". IEEE Engineering in Medicine and Biology, Vol 14, No 3, (1995) pp 292–300
9. Harris, S. J., Jakopec, M., Cobb, J., Hibberd, R. D. and Davies, B. L.: "Interactive Preoperative Selection of Cutting Constraints, and Interactive Force Controlled Knee Surgery by a Surgical Robot". Proc. of MICCAI'98, Cambridge, MA, USA, (1998) pp 996–1006
10. Harris, S. J., Jakopec, M., Cobb, J. and Davies, B. L.: "Intra-operative Application of a Robotic Knee Surgery System". Proc. of MICCAI'99, Cambridge, UK, (1999) pp 1116–1124
11. Besl, P. J. and McKay, N. D.: "A Method for Registration of 3-D Shapes". IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 14, No 2, (1992) pp 239–256
12. Zhang, Z.: "Iterative Point Matching for Registration of Free-Form Curves and Surfaces", International Journal of Computer Vision, Vol 13, No 2, (1994) pp 119–152

An Adaptive Enhancement Method for Ultrasound Images

Jun Xie, Yifeng Jiang, and Hung-tat Tsui

Electronic Engineering Department
The Chinese University of Hong Kong
{jxie,yfjiang,httsui}@ee.cuhk.edu.hk

Abstract. In this paper, we present a novel adaptive method for ultrasound (US) image enhancement. It is based on a new measure of perceptual saliency and the view-dependent feature on US images. By computing this measure on an US image, speckle noise is reduced and perceptual salient boundaries of organs are enhanced. Because of the curvature gradient based saliency measure, this method can enhance more types of salient structures than the well-known saliency network method. Meanwhile, the proposed method does not depend on the closure measure. This makes it more appropriate to enhance real images than other existing methods. Moreover, the local analysis of speckle patterns leads a good performance in speckle reduction for US images. Experimental results show the proposed enhancement approach can provide a good assistant for US image segmentation and image-guided diagnosis.

1 Introduction

US imaging allows faster and more accurate procedures due to its realtime capabilities. Moreover, it is inexpensive and easy to use. The accurate detection of organs or objects from US images plays a key role in many applications such as the accurate placement of the needles in biopsy, the assignment of the appropriate therapy in cancer treatment, and the measurement of the prostate gland volume. However, US images are commonly low-contrast, 'noisy' and with weak boundaries. Many traditional methods of image processing fail due to the speckle noise produced by the physical mechanism of ultrasonic devices.

According to the Gestalt approach [1], we perceive objects as well-organized patterns rather than separate component parts. The focal point of Gestalt theory is the idea of "grouping" or how we tend to interpret a visual field or problem in a certain way. There are four major factors that determine grouping: proximity, similarity, closure and simplicity. The objective of this work is to improve the visual effect of US images by using those grouping factors to enhance perceptual salient boundaries of organs and reduce speckle noise.

Shashua and Ullman [2] proposed the saliency network (SN) method to extract salient structures from a line drawing. The proposed saliency measure favors long, smooth curves containing only a few short gaps. They defined the "salient map" of an image as a feature to guide the grouping process. This map

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 38–45, 2004.

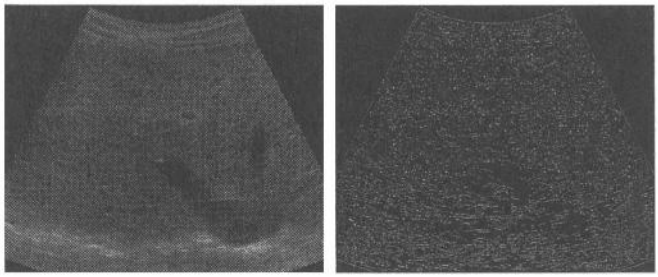


Fig. 1. A typical US image of a normal liver and its canny edge map. **Left:** The original US image. **Right:** The edge map derived via the canny operator.

was computed by an incremental optimization scheme called saliency network. This approach generally prefers long and smooth curves over short or wiggly ones. Moreover, while computing saliency, it also fills in gaps and tolerates noise.

There have been many other studies [3,4,5,6,7] on the extraction of perception structures in the last decade. Most of these existing methods, except the SN, focused on closure criteria. However, closed boundaries and curves are not general in many types of real images such as natural images and medical images. Thus, to enhance the visual effects of those images, the closure criteria should not be used. On the other hand, because the SN method prefers long curves with low total curvature, it actually treats each salient structure differently and enhance straight lines more than other salient structures.

In this paper, we introduce an adaptive speckle reduction method for US images. It is based on a salient structures extraction method [8] which favors smooth, long boundaries with constant curvatures. By combining a local speckle analysis, the proposed method can restrain the view-dependent speckle noise in US images. Experimental results of the proposed method are presented and discussed in comparison with the classical SN method and other traditional methods.

2 View-Dependent Speckle

The basis of ultrasound imaging is the transmission of high frequency sound into the body followed by the reception, processing, and parametric display of echoes returning from structures and tissues within the body. The scattering or reflection of acoustic waves arises from inhomogeneities in the tissues' density and/or compressibility. Scattering refers to the interaction between sound waves and particles that are much smaller than the sound's wavelength λ, which often occurs inside the organ. Reflection refers to such interaction with particles or objects larger than λ, which often occurs on the boundaries of organs. The typical type of scattering in the bodies arises from a diffuse population of sub-resolution particles where the arrangement of these particles is spatially random,

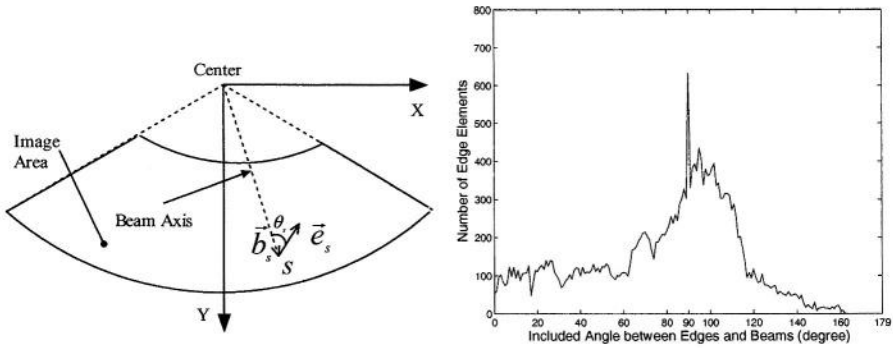


Fig. 2. An illustration of the relationship between boundaries and beam axis in US images. **Left:** Schematic explanation for the included angle θ between an element and its corresponding beam axis. **Right:** The histogram of the included angle between edges and beam directions.

and results in the speckle patterns in the image. This kind of speckle pattern can be called speckle noise since it doesn't directly reveal the physical structures under it, but the constructive and destructive interference of scattered signals from all the small structures.

Since B-mode ultrasound is most sensitive to the surfaces of structures normal to the beam, scattering shows obvious view-dependent property and its appearances in the image for a certain patch of tissues, are variant to the relative positions of the ultrasound transducer. In fact, the speckle patterns in images are observed to be elongated in the direction perpendicular to the wave beam and thus variant to the direction of beam.

Speckle occurs especially in images of the liver and kidney whose underlying structures are too small to be resolved by large wavelength ultrasound [9,10]. Fig. 1(a) shows a liver US image of a healthy human. This image has a grainy appearance, and not a homogeneous gray or black level as might be expected from homogeneous liver tissues. Fig. 1(b) is its edge map derived via the canny edge operator. It is noticed that almost all salient boundaries are perpendicular to the wave beam. Fig. 2(b) is the histogram of angles between edge fragments and the corresponding beam axis as illustrated in Fig. 2(a). These statistics confirm the assumption that most boundaries in US images are oriented nearly perpendicular to the beam directions. Based on this observation, we propose the following method to enhance US images by reducing such view-dependent speckle noise while enhancing the salient boundaries of tissues .

3 Salient Structure Extraction

In the classical SN method [2], an orientation element was defined as a vector on the image plane and the saliency of element p was defined to be the maximum

saliency over all curves emanating from p. This provides the salient map which is useful for edge enhancement. Many researchers [11,12] have explored the effects of the SN method. Wang and Li [13] showed that the salient map is a better force field for snakes since the speckle noise is reduced and the boundaries are enhanced. However, in the SN scheme, the total curvature term reduces the contribution of elements according to the accumulated square curvature from the beginning of the curve. This property makes the method favor curves with low total curvature, particularly straight lines. Thus the classical SN method is not proper for enhancing real US images that contain multiple objects with mild curving boundaries.

3.1 Curvature Gradient Based Saliency Measure

In our saliency measure, for each orientation element p on an image, the L level saliency $\Phi_L(p)$ is defined to be the maximum saliency over all curves of length L emanating from p. The saliency of a curve $\Gamma(s)$ of length l (s denotes arc length, $0 \leq s \leq l$) is defined as

$$S(\Gamma) = \int_0^l \rho^s g(s) v(s) C(s) ds \tag{1}$$

where ρ is a constant indicating the interaction level of neighboring arcs and $C(s)$ is the curvature gradient function defined as

$$C(s) = \exp^{-|\nabla k(s)|} \tag{2}$$

with $k(s)$ refers to the curvature of arc $\Gamma(s)$. Term $g(s)$ is the normalized intensity gradient of arc $\Gamma(s)$. In the SN scheme, the expansion of saliency is based on the number of non-active elements determined via a threshold of intensity gradient. This approach is not adequate for realistic applications as the selection of the appropriate threshold before a good segmentation is a difficult task. Therefore, the normalized intensity gradient is used in our measure to guide the expansion of saliency so that elements with high intensity gradient can spread their saliency farther.

Term $v(s)$ is the view-dependent factor of $\Gamma(s)$ defined as $v(s) = \exp^{|cos(\theta_s)|}$ with

$$cos(\theta_s) = \frac{\vec{\mathbf{e_s}} \cdot \vec{\mathbf{b_s}}}{|\vec{\mathbf{e_s}}||\vec{\mathbf{b_s}}|} \tag{3}$$

Vector $\vec{\mathbf{e_s}}$ denotes the orientation element at position s and $\vec{\mathbf{b_s}}$ refers to the beam vector cross $\Gamma(s)$ as illustrated in Fig. 2(a). Because most erroneous boundaries of US speckle noise are perpendicular to the beam axis as analyzed in Section 2, function $v(s)$ is introduced in our saliency measure as a perpendicularity penalty function. If element $\vec{\mathbf{e_s}}$ is perpendicular to the local beam axis, i.e. $|\cos(\theta_s)| = 0$, it will be thought to be a lower salient element for it has a high possibility to be along an erroneous speckle boundary.

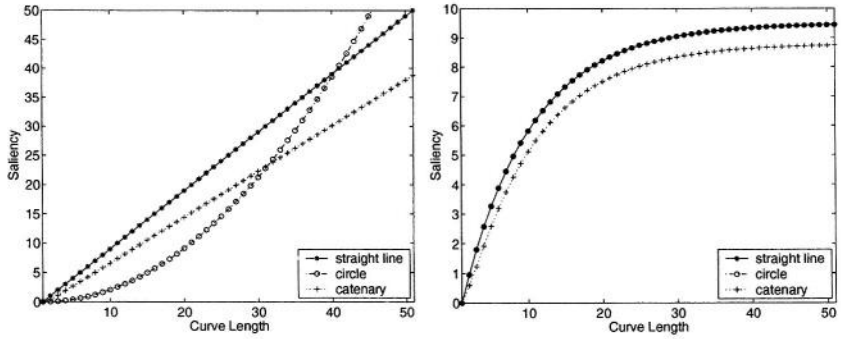


Fig. 3. Performances on three simple curves. **Left:** The saliency values calculated via the SN measure corresponding to $N = l$. **Right:** The saliency values calculated via the proposed saliency measure corresponding to $\rho = 0.9$. Because the straight line has the same saliency value as the circle, their saliency traces overlap in this diagram.

3.2 Properties of the Curvature Gradient Measure

To compare the properties of our saliency measure with the SN measure in a quantitative manner, we analyze the behavior of them by applying to three simple curves: a straight line, a circle and a catenary. First we compute the saliency of these curves using the SN measure. We consider only curves with no gaps such that $\sigma(s) = 1$ and $\rho(0, s) = 1$ for all s. Because a straight line has a constant curvature $k(s) = 0$, the saliency calculated using SN measure is $\Phi_{sn}(\Gamma) = l$. For a circle of perimeter l, the curvature is a constant $k(s) = 2\pi/l$. Therefore, the saliency of this circle is $\Phi_{sn}(\Gamma) = l^2/4\pi^2$ as derived in [12]. For a catenary $\Gamma : y = cosh(x)$, its cesàro equation is $k(s) = 1/(1 + s^2)$. Using a continuous formulation, the curvature term of the SN measure can be derived as

$$C(0, s) = \exp^{-\int_0^s k^2(t)dt} = \exp^{-\int_0^s (\frac{1}{1+t^2})^2 dt} = \exp^{-\frac{arctan(s)s^2+s+arctan(s)}{2(1+s^2)}}. \quad (4)$$

Then the saliency of this curve is computed by $\Phi_{sn}(\Gamma) = \int_0^l C(0, s)ds$. The calculated saliency values of these curves at different lengths are shown in Fig. 3(a) from which we can observe that when the lengths of these curves are short ($l < 40$), the straight line is found more salient than the other two because its sum of curvature is zero and its saliency grows linearly with the length of the line. This is consistent with the analysis that the SN measure favors straight lines more than other curves. However, when at longer lengths ($l \geq 40$), the preference of the SN measure changes and the circle is calculated as the most salient curve among these three ones. This shows that the SN measure lacks scale invariance. The above observations indicate that the SN method treats these curves differently even if they are scaled uniformly.

When using our measure, we set ρ to 0.9. For simplicity, the intensity gradient term $g(s)$ and view-dependent term $v(s)$ are both supposed to be 1 for all these

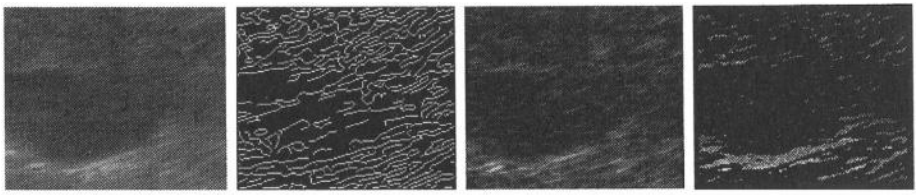


Fig. 4. An experiment on a liver US image. (a) The test image. (b) Edge map obtained via canny operator. (c) The salient map of SN method. (d) The result of the proposed method.

experiments. Therefore, our saliency equation Eq. (1) becomes

$$S(\Gamma) = \int_0^l \rho^s C(s) ds \tag{5}$$

For the straight line, the curvature gradient $C(s) = 1$ as the curvature is 0 for all arcs. Hence the saliency of a straight line of length l is

$$S(\Gamma) = \int_0^l \rho^s ds = \frac{\rho^l - 1}{\ln \rho} \tag{6}$$

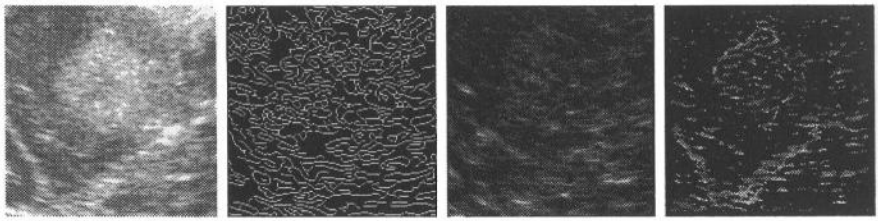
Because a circle has a constant curvature too, the curvature gradient is also equal to 1. Thus, the saliency of the circle with perimeter l is similar to that of the straight line, i.e. $S(\Gamma) = (\rho^l - 1)/\ln \rho$. Then we consider the catenary. Its curvature gradient is calculated as

$$C(s) = \exp^{-|\nabla k(s)|} = \exp^{-\frac{2s}{(1+s^2)^2}} \tag{7}$$

Therefore the saliency is

$$S(\Gamma) = \int_0^l \rho^s C(s) ds = \int_0^l \rho^s \exp^{-\frac{2s}{(1+s^2)^2}} ds \tag{8}$$

From the calculated saliency values shown in Fig. 3(b), we see that, as long as they are at a same length (short or long), the straight line has the same saliency as the circle. This satisfies our purpose that all salient structures with the same length and smoothness should obtain the same saliency. Meanwhile, it is noticed that the catenary is always less salient than the straight line and the circle. The longer their lengths are, the more different are their saliency values. The reason for this is that the catenary has higher curvature gradients than the other two curves and our proposed measure favors long curves with low curvature change rate. This novel saliency measure is computed via a proposed local search algorithm developed in [8].



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

Fig. 5. An experiment on a brain US image. (a) The original US image. (b) The edge map obtained via the canny edge operator. (c) The salient map of the SN method. (d) The result of the proposed method.

4 Experimental Results

Before applying the proposed method to real US data, the characteristics of the algorithm was first studied on several synthetic data as shown in [8]. Experimental results on real US data are shown in Fig. 4 and 5. The original test images are shown in Fig. 4(a) and 5(a). The edge images derived by the canny operator are shown in Fig. 4(b) and 5(b). They are undesirable for there are too many erroneous edges due to the speckle noise in the test US images. Fig. 4(c) and 5(c) are the salient maps of the SN method. They are better than canny edge maps for only long, smooth and strong boundaries were enhanced. The results of our method are shown in Fig. 4(d) and 5(d). Compared to results of SN method, the results of the proposed method had cleaner and thinner boundaries. This is because our local curvature gradient measure avoided the influence of noise more effectively than the curve accumulation measure. The local noise cannot be spread to the whole curve. Fig. 6 shows two experiments on the US image segmentation. We can observe that the preprocessing procedure has improved the segmentation results effectively.

5 Conclusions

In this paper, based on a novel saliency measure, we have presented an adaptive enhancement method for US images. One advantage of this method is that it can reduce the speckle noise on US images adaptively by analyzing the local speckle structures. Meanwhile, perceptual salient boundaries of organs are enhanced via computing the proposed measure of perceptual saliency. Because of this curvature gradient based saliency measure, our method can extract long, smooth and unclosed boundaries and enhance more types of salient structures equally than the SN method. Experiments show the proposed approach works very well on the given image set. This is useful for image guided surgery and other computer vision applications such as image segmentation and registration. Further research is needed to increase the searching speed of the proposed algorithm for real-time applications.

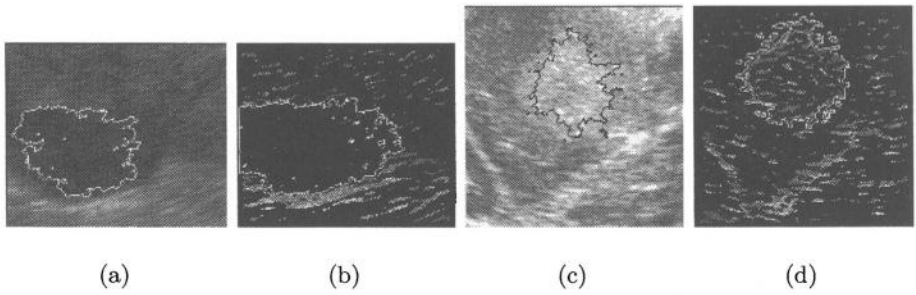


Fig. 6. Two experiments on US images using the classical Level Set Method. (a) The segmentation on an original US image. (b) The segmentation result using the enhanced US image. (c) The segmentation on another original US image. (d) The segmentation result using the corresponding enhanced US image.

References

1. M. Wertheimer. Untersuchungen zur lehre der gestalt. *Psychologische Forschung,* 4:301–350, 1923.
2. A. Sha'ashua and S. Ullman. Structural saliency: The detection of globally salient structures using a locally connected network. In *ICCV,* pages 321–327, 1988.
3. L. Herault and R. Horaud. Figure-ground discrimination: A combinational optimization approach. *IEEE Trans. on PAMI,* 15, 1993.
4. G. Guy and G. Medioni. Inferring global perceptual contours from local features. In *DARPA Image Understanding workshop,* pages 881–892, 1993.
5. S. Sharkar and K. Boyer. Quantitative measures of change based on feature organization: Eigenvalues and eigenvectors. In *CVPR,* pages 478–483, 1996.
6. S. Mahamud, L.R. Williams, K.K. Thornber, and Kanglin Xu. Segmentation of multiple salient closed contours from real images. *IEEE Trans. on PAMI,* 25(4):433–444, April 2003.
7. A. Desolneux, L. Moisan, and J.M. Morel. A grouping principle and four applications. *IEEE Trans. on Pattern Analysis and Machine Intelligence,* 25(4):508–513, April 2003.
8. Jun Xie, H.T. Tsui, and Tze Kin Lau. Edge enhancement based on salient structure extration. In *Asian Conference on Computer Vision,* pages 1152–1157, 2004.
9. C.F. Schueler, H.L. Lee, and G. Wade. Fundamentals of digital ultrasonic imaging. *IEEE Trans. Sonics Ultrasound,* 1984.
10. T.A. Tuthill, R.H. Sperry, and K.J. Parker. Deviations from rayleigh statistics in ultrasonic speckle. *Ultrason. Imaging,* 10, 1988.
11. J.S. Vilanova and K.K. Sung. Multi-scale vectorridge-detection for perceptual organization without edges. *MIT A.I. Memo,* 1992.
12. T.D. Alter and R. Basri. Extracting salient curves from images: An analysis of the salient network. *Int. Journal of Computer Vision,* 27(1):51–69, June 1998.
13. J.K. Wang and X.B. Li. A system for segmenting ultrasound images. In *ICPR,* pages 456–461, 1998.

State Space Strategies for Estimation of Activity Map in PET Imaging

Yi Tian[1], Huafeng Liu[1], and Pengcheng Shi[2]

[1] State Key Laboratory of Modern Optical Instrumentation
Zhejiang University, Hangzhou, China
[2] Department of Electrical and Electronic Engineering
Hong Kong University of Science and Technology, Hong Kong
{eeliuhf, eeship}@ust.hk

Abstract. In this paper, we explore the usage of state space principles for the estimation of activity map in tomographic PET imaging. The proposed strategy formulates the dynamic changes of the organ activity distribution through state space evolution equations and the photon-counting measurements through observation equations, thus makes it possible to unify the dynamic reconstruction problem and static reconstruction problem into a general framework. Further, it coherently treats the uncertainties of the statistical model of the imaging system and the noisy nature of measurement data. The state-space reconstruction problem is solved by both the popular but suboptimal Kalman filter (KF) and the robust H_∞ estimator. Since the H_∞ filter seeks the minimum-maximum-error estimates without any assumptions on the system and data noise statistics, it is particular suited for PET imaging where the measurement data is known to be Poisson distributed. The proposed framework is evaluated using Shepp-Logan simulated phantom data and compared to standard methods with favorable results.

1 Introduction

Accurate and fast image reconstruction is the ultimate goal for many medical imaging modalities such as positron emission tomography (PET). Accordingly, there have been abundant efforts devoted to tomographic image reconstruction for the past thirty years. In PET imaging, the traditional approach is based on the deterministic filtered backprojection (FBP) method [3]. Typical FBP algorithms do not, however, produce high quality reconstructed images because of their inability to handle the Poisson statistics of the measurements, i.e. the counts of the detected photons. More recently, iterative statistical methods have been proposed and adopted with various objective functions, including notable examples such as maximum likelihood (ML) [12], expectation maximization (EM) [9], ordered-subset EM [2], *maximum a posteriori* (MAP) [10], and penalized weighted lease-squares (PWLS) [4].

For any statistical image reconstruction framework, one must consider two important aspects of the problem: the statistical model of the imaging system

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 46–53, 2004.

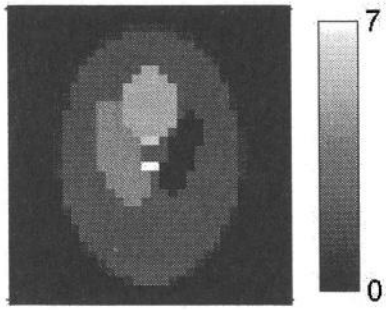


Fig. 1. Digital Shepp-Logan phantom used in the experiments (left) and scale map (right)

and the noisy nature of measurement data. It is clear that an accurate statistical model is a prerequisite for a good reconstruction [1]. However, all aforementioned existing works do not consider the uncertainties of the statistical model, while in practical situations it is almost impossible to have the exact model information. In addition, these existing methods assume that the properties of the organs being imaged do not change over time. In the dynamic PET, however, the activity distribution is time-varying and the goal is actually to obtain dynamic changes of the tissue activities [8].

In this paper, we present a general PET reconstruction paradigm which is based on the state space principles. Compared to earlier works, our effort has three significant novel aspects. First, our approach undertakes the uncertainties of both the statistical model of the imaging system and the measurement data. Secondly, our method formulates the dynamic changes of the organ activity distribution as state space variable evolution, thus makes it possible to unify the dynamic reconstruction problem and static reconstruction problem into a general framework. Finally, two solutions are proposed for the state space framework: the Kalman filtering (KF) solution which adopts the minimum-mean-square-error criteria, and the H_∞ filter which seeks the minimum-maximum-error estimates. Since the H_∞ principle makes no assumptions on the noise statistics, it is particular suited for PET imaging where the measurement data is known to be Poisson distributed. An evaluation study by using Shepp-Logan simulated phantom data of our proposed strategy is described. Experimental results, conclusions and future work are also presented.

2 Methodology

2.1 State Space Representation

In emission tomography such as PET, the goal is to reconstruct the emission/radioactivity distribution x from the projected measurement data y (the

photon counts). In general, the measurement of the emission scan can be described by the measurement or observation equation:

$$y = Dx + e \tag{1}$$

where x is a N_p x 1 vector and N_p is the total number of image voxels, y is a N_y x 1 vector and N_y is the total number of the detector pairs in the projection measurement. D is a N_yxN_p detection probability matrix with it $(i,j)^{th}$ element value equals to the probability of detecting an event from the j^{th} voxel recorded by the i^{th} detector pair, with consideration of the detector efficiency and the photon attenuation. In addition, the measurement noise e models the uncertainty of the measurement data, which mainly accounts for the presence of the scattered and random events in the data, and we assume $E[e(t)e(s)] = R_e\delta_{ts}$. We aim to use the noisy observations y to recover the distribution x in some optimal sense.

The state equation of the imaging system, which describes the radioactivity of the pixels, can be written in the form of

$$x(t+1) = Ax(t) + v \tag{2}$$

with some initial activity x_0. The system noise v models the statistical uncertainty of the imaging model, with $E[v(t)v(s)] = Q_v\delta_{ts}$. In general, Eqn. 2 represents the dynamic changes of the state variable x, and it reduces to the conventional static reconstruction problem when the transition matrix A is an identity matrix.

2.2 Kalman Filer Solution

The Kalman filter adopts a form of feedback control in estimation: the filter estimates the process state at some time and then obtains the feedback in the form of (noisy) measurements [6]. Hence, the time update equations of the Kalman filter are responsible for projecting forward (in time) the current state and error covariance estimates to obtain the *a priori* estimates for the next time step, while the measurement update equations are responsible for the feedback – i.e. for incorporating a new measurement into the *a priori* estimate to obtain an improved *a posteriori* estimate. And the final estimation algorithm resembles that of a predictor-corrector algorithm for solving numerical problems.

A recursive procedure is used to perform the state estimation of Equations (2) and (1) [5]:

1. Initial estimates for state $\hat{x}_0$ and error covariance $P(0)$.
2. Time update equations, *the predictions,* for the state

$$\hat{x}^-(t) = A\hat{x}(t-1) \tag{3}$$

and the error covariance

$$P^-(t) = AP(t-1)A^T + Q_v(t) \tag{4}$$

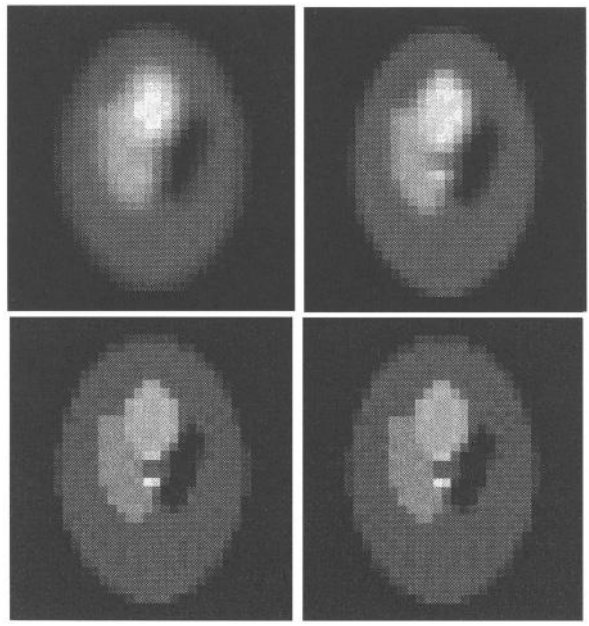


Fig. 2. Reconstruction of Shepp-Logan phantom using noise-free measurement. From left to right. Top: FBP results, ML-EM results. Bottom: KF results and H_∞ results.

3. Measurement update equations, *the corrections,* for the Kalman gain

$$L(t) = P^-(t)D^T(DP^-(t)D^T + R_e(t))^{-1} \tag{5}$$

the state

$$\hat{x}(t) = \hat{x}^-(t) + L(t)(y - D\hat{x}^-(t)) \tag{6}$$

and the error covariance

$$P(t) = P^-(t) - L(t)(DP^-(t)D^T + R_e(t))L^T(t) \tag{7}$$

2.3 $H\infty$ Solution

The Kalman filter requires the prior knowledge of the statistical properties of the Gaussian state variables and noises, which may not agree with the noise nature of PET measurement. Meanwhile, the mini-max H_∞ strategy does not impose such restrictions and only makes assumptions on the finite disturbance energy. It is thus more robust and less sensitive to noise variations and modelling assumptions [11].

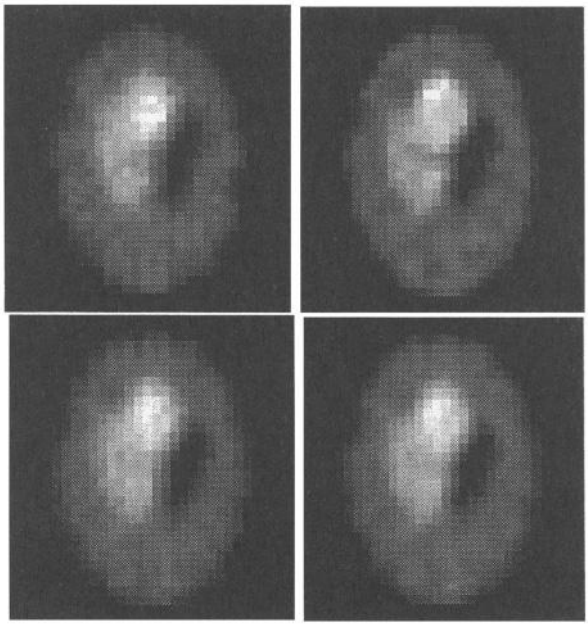


Fig. 3. Reconstruction of Shepp-Logan phantom using $SNR = 15dB$ input data. From left to right. Top: FBP results, ML-EM results. Bottom: KF results and H_∞ results.

Along with the state and measurement equations, the required output equation of the H_∞ filter is constructed as:

$$z(t) = Fx(t) \tag{8}$$

where the output variable $z(t)$ is the linear combination of the kinematics states $x(t)$, and the entries of the known output matrix F are specified by the user. In our case, F is just an identity matrix.

We adopt the H_∞ filtering formulation of [11], which has a similar structure to the Kalman filter but with different optimizing criteria. While the Kalman filter calculates the estimation error using the H_2 norm and minimizing the mean-square error, the H_∞ filter evaluates the error in terms of H_∞ norm through the performance measure:

$$J = \frac{\sum_{t=0}^{N-1} \|z(t) - \hat{z}(t)\|^2_{Q(t)}}{\|x_o - \hat{x}_o\|^2_{p_o^{-1}} + \sum_{t=0}^{N-1} (\|v(t)\|^2_{N(t)^{-1}} + \|e(t)\|^2_{V(t)^{-1}})} \tag{9}$$

with $N(t)$, $V(t)$, $Q(t)$ and p_o the weighting matrices for the process noise, the measurement noise, the estimation error, and the initial conditions respectively, and $\hat{x}_o$ and $\hat{z}_o$ the *a priori* estimates of x_o and z_o. The denominator of J can be regarded as the energy of the unknown disturbances, while the numerator is the

Inputs	Methods	mean ± std
Noise-free	FBP	0.0805 ± 0.4081
	ML-EM	$6.43\text{x}10^{-7}$ ± 0.3061
	KF	$1.81\text{x}10^{-7}$ ± 0.1471
	H_∞	$1.26\text{x}10^{-7}$ ± 0.1445
SNR=15dB	FBP	0.0905 ± 0.4263
	ML-EM	0.0046 ± 0.3361
	KF	0.076 ± 0.42
	H_∞	0.0085 ± 0.2479

Table 1. Comparative studies of estimated activity distribution. Each data cell represents reconstruction error: the mean ± standard derivation.

energy of the estimation error. Obviously, a desirable estimator is the one for which this energy gain is small so that the disturbances are attenuated. Hence, the H_∞ filter aims to choose the estimator such that the worst-case energy gain is bounded by a prescribed value:

$$\sup J \; < \; 1/\gamma \tag{10}$$

where *sup* means the supremum and $1/\gamma$ is the noise attenuation level. The robustness of the H_∞ estimator arises from the fact that it yields an energy gain less than $1/\gamma$ for all bounded energy disturbances no matter what they are.

We have adopted a game theoretic algorithm which can be implemented through recursive updating of the filter gain $K(t)$, the Riccati difference equation solution $P(t)$, and the state estimates $\hat{x}(t)$:

$$K(t) = AP(t)S(t)D^T V(t)^{-1} \tag{11}$$

$$P(t+1) = AP(t)S(t)A^T + N(t) \tag{12}$$

$$\hat{x}(t+1) = A\hat{x}(t) + K(t)(y - D\hat{x}(t)) \tag{13}$$

$$S(t) = (I - \gamma \bar{Q}(t)P(t) + D^T V(t)^{-1} DP(t))^{-1}$$

$$\bar{Q}(t) = F^T Q(t) F$$

It should be also noted that the weighting parameters $(N(t), V(t), Q(t), p_o)$ and the performance bound (γ) should be carefully adjusted in order to fulfil the performance criteria in (9).

3 Experiments and Discussion

An evaluation of the proposed frameworks is conducted to verify its capability to faithfully and robustly reconstruct the attenuation map from the noisefree/noisy measurements of the known activity distribution. An elliptical object is simulated, with a constant activity and a Shepp-Logan phantom type attenuation distribution, as shown in Fig 1 [7]. The sinograms has 34 radial bins and 120

angles uniformly sampled over 180 degrees, which is labeled as the ideal measurement data. $SNR = 15dB$ level of Poisson noise is then added to generate the noisy data. With the state space algorithms using Kalman and H_∞ filters, the activity distribution is reconstructed from the simulated clean and noisy sinogram. In our current implementation, the process noise covariance matrix Q_v, the measurement noise covariance matrix R_e, and the H_∞ weighting parameters $N(t)$, $V(t)$, $Q(t)$, and p_o are all set as diagonal matrices, and these values are fixed during the estimation process. Experiments are also conducted using the standard FBP and ML-EM methods [12] for comparison.

Fig.2 shows the estimated activity maps from the FBP, ML-EM, KF, and H_∞ algorithms, in which the initial values of the radioactivity distribution x_0 are assigned to an identity matrix and the measurement data is noise-free. The estimated maps using the $SNR = 15dB$ noisy data are shown in Fig. 3.

A detailed statistical analysis on the estimation results against the ground truth phantom map is performed. Let $\hat{x}$ be final reconstruction results and x_{tr} be the ground truth, we have the following error definitions:

$$mean = \frac{1}{N_p}\sum(\hat{x} - x_{tr}) \tag{14}$$

$$std = \left(\frac{1}{N_p - 1}\sum(\hat{x} - x_{tr})^2\right)^{0.5} \tag{15}$$

The analysis results are summarized in Table 1. For the noise-free data, the H_∞ algorithm gives the best result among the four strategies. For the noisy data, however, the ML-EM gives result with somewhat smaller mean error. Nevertheless, it is clear that the H_∞ results have smaller, and thus desired, variances for both clean and noisy data.

We want to point out that, in our current implementation for noisy cases, the noise covariance matrices Q_v and R_e (KF framework), as well as the weighting parameters and the performance bound γ (H_∞ framework), are set to some empirically fixed values, which are not optimal. Ideally, these parameters should be adaptively updated during the estimation process, similar to the ML-EM algorithm, especially for realistic problems. While improvement of the results are expected from such procedure, detailed investigation on this issue is still underway.

4 Conclusion

In this paper, we have presented a state space framework for estimation activity maps from tomographic PET measurement data. Our approach adopts a robust system identification paradigm, and is derived and extended from the KF and H_∞ filtering principles, where the latter is particularly suitable for PET imaging reconstruction where the measurement data is known to be Poisson distributed. Analysis and experiment results with Shepp-Logan simulation phantom data demonstrate the power of the new proposed method.

Acknowledgments This work is supported by the Hong Kong Research Grant Council for HKUST6151/03E and by the National Basic Research Program of China (No: 2003CB716104).

References

[1] C. Comtat, P.E. Kinahan, M. Defrise, C. Michel, and D.W. Townsend. Fast reconstruction of 3d pet data with accurate statistical modeling. *IEEE Trans. Nucl. Sci.,* 45:1083–1089, 1998.

[2] H. Erdogan and J.A. Fessler. Ordered subsets algorithms for transimission tomography. *Phys. Med. Biol.,* 44(11):2835–2851, 1999.

[3] J. A. Fessler. Statistical image reconstruction methods for transmission tomography. *SPIE Handbook of Medical Imaging.*

[4] J. A. Fessler. Penalized weighted least-squares image reconstruction for positron emission tomography. *IEEE Transactions on Medical Imaging,* 13(2):290–300, 1994.

[5] A. Gelb. *Applied Optimal Estimation.* MIT Press, Cambridge, MA, 1974.

[6] T. Glad and L. Ljung. *Control Theory.* Taylor & Francis, London, 2000.

[7] A. C. Kak and Malcolm Slaney. Principles of computerized tomographic imaging. *IEEE Press,* 1988.

[8] C.M. Kao, J.T. Yap, J. Mukherjee, and M.N. Wernick. Image reconstruction for dynamic pet based on low-order approximation and restoration of the sinogram. *IEEE Transactions on Medical Imaging,* 16:738–749, 1997.

[9] K. Lange and R. Carson. Em reconstruction algorithms for emission and transmission tomography. *J. Comput. Asst. Tomogr.,* 8:306–316, 1984.

[10] J. Nuyts, D. Beque, P. Dupont, and J. Mortelmans. A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography. *IEEE Trans. Nucl. Sci.,* 49:56–60, 2002.

[11] X. Shen and L. Deng. A dynamic system approach to speech enhancement using the $\mathcal{H}_\infty$ filtering algorithm. *IEEE Trans. Speech and Audio Processing,* 7(4):391–399, 1999.

[12] L.A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Medical Imaging,* 1:113–122, 1982.

Applying ICA Mixture Analysis for Segmenting Liver from Multi-phase Abdominal CT Images

Xuebin Hu[†], Akinobu Shimizu[†], Hidefumi Kobatake[†], Shigeru Nawano[‡]

† Graduate School of Bio-Applications and Systems Engineering,
Tokyo University of Agriculture & Technology
2-24-16 Naka-cho, Koganei-shi, Tokyo, 184-8588 Japan
{huxb, simiz, kobatake}@cc.tuat.ac.jp
‡ National Cancer Center Hospital East
6-5-1 Kasiwanoha, Kasiwa-shi, Chiba, 277-8577 Japan
snawano@east.ncc.go.jp

Abstract. Liver segmentation is an important task in the development of computer-aided multi-phase CT diagnosis system of liver. This paper presents a new approach for segmenting liver from multi-phase abdominal CT images using ICA mixture analysis. In particular, we use the variational Bayesian mixture of ICA method [1] to analyze three-dimensional four-phase abdominal CT images. The analysis results show that the CT images could be divided into a set of clinically and anatomically meaningful components. As to our concern, the organs that surround the liver and have similar intensities, such as stomach, kidney, are nearly completely separated from the liver, which makes the segmentation become much easier than on the original CT images.

1 Introduction

The rapid development of medical imaging technologies allow physicians to glean potentially life-saving information by peering non-invasively into the human body. At the same time, several hundreds of slice images per patient become a heavy burden for physicians to efficiently grasp all the useful information, especially in multi-phase CT diagnosis. The development of medical imaging technology is rapidly going towards more advanced computer-aided systems.

Liver segmentation is an important task in the development of the computer-aided diagnosis system for multi-phase CT diagnosis. Although modern imaging devices provide exceptional views of internal anatomic structures, the use of computers to quantify and analyze the embedded structures with accuracy and efficiency is still limited. The difficulty in liver segmentation comes from the fact that the surrounding tissues, such as stomach and muscle, have similar CT values and sometimes contact with each other, which may cause the boundaries of liver to be indistinct and disconnected. Traditional low-level image processing techniques that consider only local information can make incorrect assumptions and generate infeasible object boundaries. On the other hand, deformable models appear to give promising results because of their ability to match the anatomic structures by exploiting constraints derived from the image together with prior knowledge about the location, size, and

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 54-61,2004.

shape of these structures. However, due to the variability and complexity of human anatomical structure, it is still difficult to achieve a satisfying result in some complicated situation.

Independent component analysis (ICA) is one of the most powerful tools in multivariate analysis. It has attracted a great of research interests in medical signal and image processing. The examples include in EEG, ECG de-noising and removing artifacts, coloring multi-channel MR imaging data, extracting blood vessel-related component from dynamic brain PET images, and fMRI data analysis, etc. In this paper, we use the recent development of ICA mixture model to analyze four-phase abdominal CT images for finding the latent meaningful components. ICA mixture model was first formulated by Lee et al. in [2], in which it is defined as that the observed data are categorized into several mutually exclusive classes, and the data in each class are modeled as being generated by a linear combination of independent sources. It relaxes the assumption of standard ICA that the sources must be independent, and shows improved performance in data classification problems [3].

Four-phase abdominal CT images are taken at different phases before and after the contrast material injected. It is often used as an effective measure for tumor detection. ICA mixture analysis results show that the CT images could be divided into a set of clinically and anatomically meaningful components. We have reported our experimental finding in [4] that the ICA mixture model could significantly increase the contrast of tumor area with its surrounding tissue so that increasing its detect-ability. In this paper, we discuss its application in three-dimensional liver segmentation. In the subsequent pages of this paper, the ICA mixture model used in this paper is briefly introduced in section 2. Section 3 presents the ICA results. In section 4, we discuss its application in liver segmentation. At the last is the conclusion.

2 ICA Mixture Model

2.1 ICA Mixture Model

Let us first recall the ICA model: $\boldsymbol{x}=\boldsymbol{A}\boldsymbol{s}$, where $\boldsymbol{x}=\{\boldsymbol{x}^n, n=1,\cdots,N\}$ is the K-dimensional vector of the observed variables, $\boldsymbol{s}=\{\boldsymbol{s}^n, n=1,\cdots,N\}$ is the L-dimensional vector of latent variables that are assumed non-Gaussian and mutually independent, and $\boldsymbol{A}$ is an $K\times L$ unknown mixing matrix. In signal processing nomenclature, K is the number of sensors and L is the number of latent sources. The observed data $\boldsymbol{x}$ is expressed as generated by a linear combination of underlying latent variables $\boldsymbol{s}$, and the task is to find out both the latent variables and the mixing process.

Comparatively, ICA mixture model is defined as that the observed data $\boldsymbol{x}$ are categorized into several mutually exclusive clusters, and the data in each cluster are modeled as being generated by a linear combination of independent sources. Suppose a C-cluster model, then the observed variables $\boldsymbol{x}$ is divided into $\{\boldsymbol{x}_c \in \boldsymbol{x}, c=1,\cdots C\}$, and the variables in each cluster $\boldsymbol{x}_c$ are generated by linearly transforming an unobserved source vector $\boldsymbol{s}_c$, of dimension L_c with added Gaussian noise $\boldsymbol{e}_c$ [1],

$$\boldsymbol{x}_c = \boldsymbol{A}_c\boldsymbol{s}_c + \boldsymbol{y}_c + \boldsymbol{e}_c \tag{1}$$

where $\boldsymbol{y}_c$ is an *K*-dimensional bias vector, and $\boldsymbol{e}_c$ is *K*-dimensional additive noise. Note in the ICA model, we supposed a noise-free model, and the bias term is also omitted because it could be easily deducted from the observed variables by removing the means. Equation (1) acts as a complete description for cluster *c* in the data density. The bias vector, $\boldsymbol{y}_c$, defines the position of the cluster in the *K*-dimensional data space, $\boldsymbol{A}_c$ describes its orientation and $\boldsymbol{s}_c$ describes the underlying manifold. The noise, $\boldsymbol{e}_c$, is assumed to be zero-mean Gaussian and isotropic [1].

In the ICA mixture model, the probability of generating a data vector $\boldsymbol{x}^n$ from a *C*-component mixture model given model assumptions *M* is described by [1]:

$$p(x^n \mid M) = \sum_{c=1}^{C} p(c \mid M_0) p(x^n \mid M_c, c) \tag{2}$$

A data vector $\boldsymbol{x}^n$ is generated by choosing one of the C components stochastically under $p(c \mid M_0)$ and then drawing from $p(x^n \mid M_c, c)$. $M = \{M_0, M_1, \cdots, M_C\}$ is the vector of component model assumption, M_c, and assumption about the mixture process, M_0. The variable *c* indicates which component of the mixture model is chosen to generate a given data vector $\boldsymbol{x}^n$. $p(c \mid M_0)$ is a vector of probabilities. $p(x^n \mid M_c, c)$ is the component densities and assumed to be non-Gaussian. The mixture density $p(x^n \mid M)$ is known as the evidence for model *M* and quantifies the likelihood of the observed data under model *M*.

2.2 Variational Bayesian Mixture of ICA

In [2], Lee et al. proposed the ICA mixture model and used the extended Infomax algorithm [5] to switch the source model between sub-Gaussian and super-Gaussian regimes, the parameters of the model were learned by using gradient ascent to maximize a log-likelihood function. The variational Bayesian mixture of ICA method (vbmoICA) used in this paper has the characteristics that are different from Lee's method in two aspects. The first is that, instead of a predefined density model, vbmoICA uses a fully adaptable mixture of Gaussians as the ICA source model, allowing complex and potentially multimodal distributions to be modeled. The second difference is that the ICA mixture model is learned under Bayesian framework, which is carried through to the mixture model. This allows model comparison, incorporation of prior knowledge, control of model complexity thus avoiding over-fitting.

In particular, a variational method [6] is used because the Bayesian inference in such a complex model is computationally intensive and intractable. Variational method is an approximation approach for converting a complex problem into a simple problem. In the vbmoICA, it approximates the posterior distributions by constructing a lower bound on it, and attempts to optimize this bound using an iterative scheme that has intriguing similarities to the standard expectation-maximization algorithm. The parameters of the model are learned by alternating between estimating the posterior distribution over hidden variables for a particular setting of the parameters and then re-estimating the best-fit parameters given that distribution over the hidden variables. Due to space limit, the details of learning rules are referred to [1].

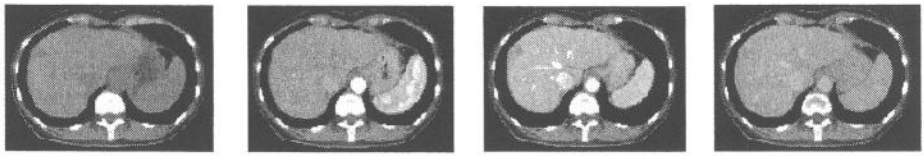


Fig. 1. An example of the four-phase CT images (the 22nd slice). From left to right they are the pre-contrast, early, portal, and late phase, respectively. The four images are displayed within the range from -110 to 190 H.U.

3 Analysis of Abdominal CT Images

3.1 Test Materials and Procedures

The test materials are a set of multi-phase CT images of human abdomen. The data were provided by the National Cancer Center Hospital East, Japan, which were collected at four phases before and after contrast material injected. They are called the pre-contrast, early, portal and late images, respectively. The dimensions are $512 \times 512 \times 154 \sim 267$, spatial resolution is $0.546 \sim 0.625$ mm, and the space between slices is 1 mm. The later three phase images were firstly registered to the early phase CT by template matching of the center of spine [7]. Then the four CT images were transformed to make each voxel to be cubic. The four images were compressed by one-third, and furthermore trimmed to $116 \times 170 \times 86$ by removing the volume that contains air only. Fig.1 shows one example of the four-phase CT images.

The structure of model could be decided by calculating a term called negative free-energy (NFE) in the vbmoICA. In the analysis, we experimentally set the number of clusters equals to four, and the number of independent components (IC) within each cluster equals to three. Because the dataset is quite big, we remove the voxels of which the CT intensities are lower than -900 H.U, which correspond to the air and some artifacts outside the body contour. The four images are then vectorised into four channels and used as the observed variables in the ICA mixture model. Eighty thousands samples are randomly drawn from the whole dataset and are used to learn the parameters of the model as well as the probability distributions of independent components. The learned model is then used to classify the whole dataset, and decompose each cluster into the corresponding independent components, which corresponds to the source vector $\boldsymbol{s}_c$ in Eq. (1). Then we put back the removed voxels accordingly and recover the derived one-dimensional ICs into three-dimensions.

3.2 ICA Analysis Results

Figure 2 shows a slice of the derived independent components, which are corresponding to the original CT images showed in Fig.1. As we assume a four-cluster model and each consists of three ICs, totally we got twelve features. The presented independent features are very meaningful. The first row of ICs (1-3) comprises a cluster in which vessels, spine, stomach and spleen are included. It is noted that the reason why spleen and stomach are not so apparent is just because we

display these features at their full intensity ranges, respectively. The second class of ICs that consists of the components numbered (4-6) appear to be responsible for generating the liver, the muscle and fat tissue at periphery of body. The third class of ICs (7-9) appear to be responsible for generating the lung, where the ribs seem to be classified into the forth cluster (10-12).

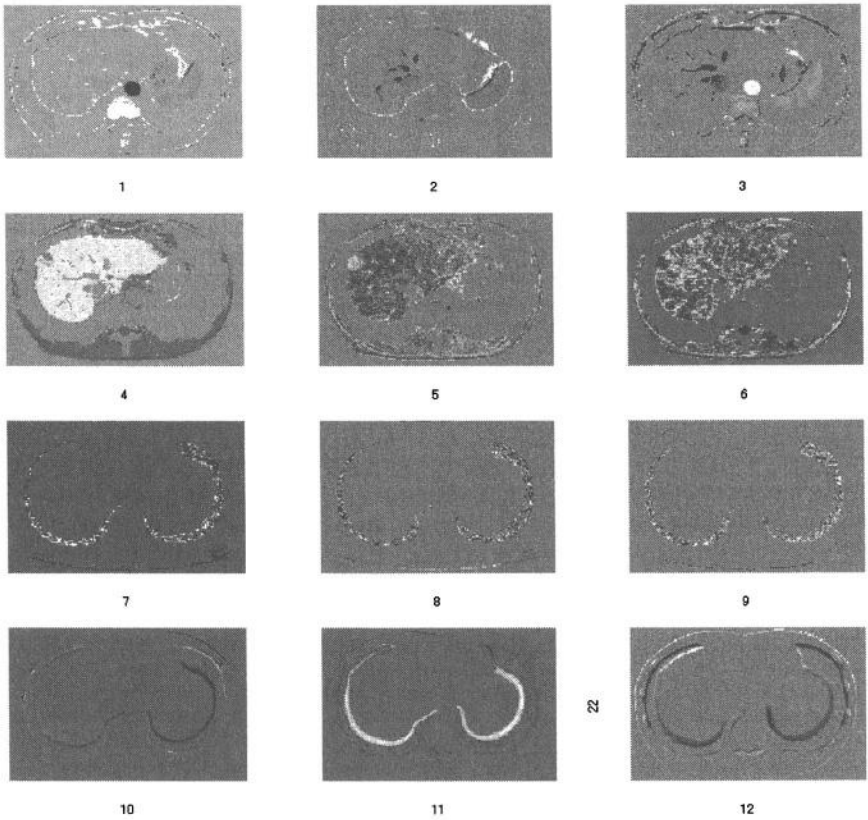


Fig. 2. A slice of the derived independent components, which corresponds to the original CT slice shown in of Fig.1. From top to down, each row consists of three ICs and comprises a cluster. These ICs are displayed at their full intensity ranges, respectively.

Because there are registration errors between the four images, the clusters have not been perfectly classified by the organ. For example, the third class of ICs actually only constitute a part of the lung, precisely speaking, they constitute the overlapped area of lung between the four-phase images. Simultaneously, the fourth cluster includes some voxels that belong to lung. The situation is similar for liver. In the preprocessing step, we simply used a template matching method for the registration. We once tried some other methods like using polynomial transform with mutual information as the metric, but achieved no better result. A good registration approach between abdominal medical images is not only important to the topic in this paper, but also important for some other purposes. It is to be discussed in next section.

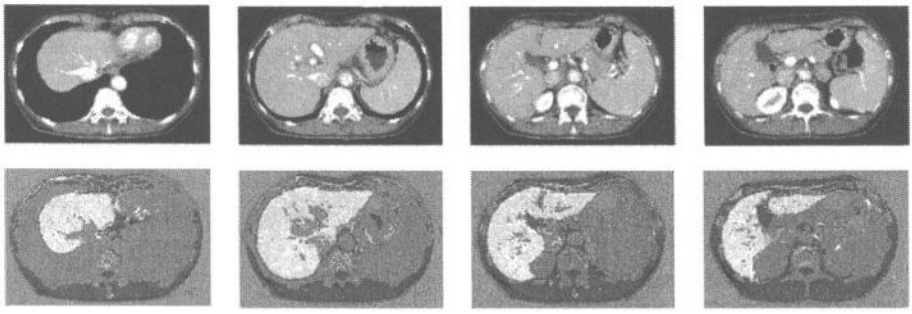


Fig. 3. Comparison of the derived independent component (lower row) with the portal phase CT image (upper row). The column drawings, from left to right, are the slices of 15^{th}, 34^{th}, 47^{th} and 55^{th}, respectively. The organs of left kidney, stomach, left lung that contact with liver are excluded in the cluster, which highlights the liver. The gallbladder is included in the cluster, but its contrast with liver has been increased.

4 Applications in Liver Segmentation

The anatomically and clinically meaningful results as shown in Fig.2 imply the prospects in various potential applications. Here we discuss its application in segmentation. Fig.3 shows several slices at different depths comparing to the corresponding portal phase CT images, respectively. As to our concern of liver segmentation, the most attractive point to us is that the organs such as stomach, lung, kidney that surround the liver are classified into other clusters, or in other word, they have been nearly completely separated from the liver. This releases the most difficult problem in liver segmentation that the surrounding tissues like stomach may contact with liver and has nearly the same intensity value, and consequently makes the segmentation becomes much easier than on the original CT images. We simply use binarization processing, and followed by the morphology operations of closing and opening, each once, to extract the liver area. Fig.4 shows the extracted boundary of liver that are displayed on the four-phase original images at two different depths, respectively. The boundary at the 47th slice (upper row in Fig.4) is extracted quite well, where the boundary at the 22nd slice (lower row in Fig.4) is under-extracted.

The problem in using ICA mixture model for liver segmentation is that it seems to only extract the overlapped area between the livers in the four images. If there exists error in registration, some voxels that are close to the boundary may be classified into the cluster that emphasizes the other organs. One example could be found in Fig.2-11, in which the arc-shaped white object in the left of the image, roughly speaking, should be a part of lung for portal phase image and be a part of liver in other three phases. In such case, segmentation may be good at a certain phase, and may be under-extracted at some other phases. The lower row of Fig.4 is just such an example. The extracted boundary segments the liver area in the portal phase well, but is smaller than the liver areas in other three phase images.

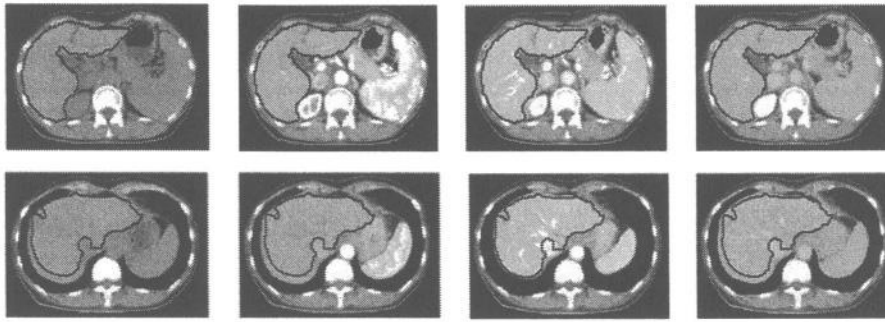


Fig. 4. Examples of segmentation result. From left to right are the pre-contrast, early, portal and later phases respectively. The upper row shows a good example at 47th slice. The lower row is under-extracted (except for the portal phase) at 22nd slice due to the registration error. Note in the lower drawings, the concavity at the left-up side of contour is result from a tumor at there, which could be seen in Fig.1 and clearly be found in the second row of Fig.2.

The reason why the extracted boundary at 47th slice is good but at 22nd slice is inadequate is thought mainly due to the fact that the variation of lung has no direct effect on the former but affects the later greatly. Although the multiple-phase abdominal CT images are taken when the patients are in still condition, the time needed is normally longer than a breath. When the patient breathes in or out, the shape as well as the pressure from lung changed, and at the same time the organs that contact with lung (e.g. liver, spleen, etc.) are deformed due to the changing pressure from the lung. The variation of lung and the consequent variation of inside organs are cooperative but in an "inverse" manner. This implies that global transform is not suitable for the registration task. On the other hand, capture range might be a problem for the registration methods using local transforms because the variation is large. Furthermore, due to the large variability and complexity of anatomic structure, it is quite difficult for the currently developed metrics to lead the transforms to a good registration. It might be an easier way to do the job in a hierarchical way that to register only the lung at the first, and use fully the information from the variation of lung into the subsequent registration of the inside organs. This will help to decrease the complexity in part, and at least be helpful to the concerned segmentation of liver.

5 Conclusions

We adopt the variational Bayesian mixture of ICA method on four sets of images, and all achieved similar meaningful results. The reason why ICA mixture model can achieve such a meaningful results, for example the cluster that emphasizes liver could exclude the stomach that has similar intensity value, is not much clear. The explanation could be given is by Choudrey et. al. in [1] that the clusters in the model are learned to be represented by their own local coordinate systems that are constructed under the independent assumption.

The anatomically meaningful features make us believe that ICA mixture model can be used as an efficient tool in various medical image processing tasks. It is proved useful in this paper to use the derived independent component for a better segmentation. As to our concern, when the breath is steady, it is readily to obtain an adequate segmentation of liver. If the variation of lung is large, there will appear some errors, especially at the upper part of liver that contact the liver. Currently, we are working on development of a hierarchical registration method that taking into consideration the variation of lung due to breath. Our subsequent works include the development of computer-aided multi-phase CT diagnosis system of liver, and the application of ICA mixture model to other medical image processing tasks are also interested to us.

Acknowledgement

This work is supported in part by the Grant-in-Aid for Scientific Research on Priority Areas: Intelligent Assistance in Diagnosis of Multi-dimensional Medical Images, from the Ministry of Education, Sports, Culture, Science and Technology, Japan.

References

[1] Choudrey R.A, Roberts S.J: Variational mixture of Bayesian independent component analyzers. Neural Computation, vol. 15(1), Jan. 2003.

[2] Lee T.W, Lewicki M.S: Upsupervised image classification, segmentation, and enhancement using ICA mixture models. IEEE Trans. on Image Processing, vol. 11, No. 3, pp. 270-279, March 2002.

[3] Lee T.W, Lewicki M.S, Sejnowski T. J: ICA mixture models for unsupervised classification and automatic context switching. Proceeding of International workshop on Independent Component Analysis, pp. 209-214, 1999.

[4] Hu X-B, Shimizu A, Kobatake H, et al: Independent component analysis of four-phase abdominal CT images. Proc. of MICCAI2004 (to appear).

[5] Lee T. W, Girolami M, Sejnowski T.J: Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian sources. Neural Computation, vol. 11, no. 2, pp. 409-433, 1999.

[6] Attias H: Learning parameters and structure of latent variable models by variational Bayes. Electronic Proceedings of the Fifteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI1999), http://www2.sis.pitt.edu/˜dsl/UAI/uai.html.

[7] Hitosugi T, Shimizu A, Tamura M, et al: Development of a liver extraction method using a level set method and its performance evaluation. Journal of Computer Aided Diagnosis of Medical Images, vol. 7, No. 4-2, Jun. 2003 (in Japanese).

Extracting Pathologic Patterns from NIR Breast Images with Digital Image Processing Techniques

Kaiyang Li[1], Yan Xiang[2], Xuandong Yang[1], and Jiming Hu[1]

[1] Lab of Biomedical Engineering, Department of Physics, Wuhan University, China,
lky@whu.edu.cn
[2] Department of Biomedical Engineering
Kunming Science and Technology University, China
yansunn_whu@sohu.com

Abstract. With a method of digital image processing including a variety of units, e.g., Laplacian filters, rank value filters, edge detectors, gain units and summation units, some pathologic patterns can be extracted from near-infrared (NIR) breast images. These pathologic patterns express the structural and the vascular information of mammary tissue. The information is concealed in primary NIR breast images, and not appear in common condition. The analysis by these pathologic patterns combined with the primary images, the possibility of false positive and false negative diagnosis is reduced, as results, and the diagnosis on breast diseases is improved obviously.

1 Introduction

The incidence of breast cancer is increasing and it will be the most common malignant tumor in the world, so diagnosis and recognition of breast cancer is so important. The first clinical result on optical transillumination imaging of the female breast has been performed as far back as 1929 [1]. The contrast can be improved by recording a photographic image of the transilluminated light at NIR wavelengths [2]. A model system named telediaphanography was described by Watmough DJ [3]. Since the resolution is very low (about 2 cm), and only very large tumors (or those near the surface) are detectable, transillumination techniques yield a low sensitivity and specificity. The low specificity results in many women without cancer to be recommended to undergo breast biopsy due to false-positive findings. Conversely, the limited sensitivity leads that some cancers are difficult or impossible to detect until the tumor is large [4].

At present, two types of the developed NIR imaging techniques have been studying on frequency-domain [5] and time-resolved [6]. Different from the two techniques, a practicable method of digital image processing to extract some specific information from primary NIR breast images was developed in this paper. Two of operators based on a method of digital image processing available for NIR breast images are represented.

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 62-69, 2004.

2 Materials and Methods

2.1 NIR Breast Imaging and Diagnosis

1466 stochastic clinical female patients were diagnosed with self-made NIR breast diagnostic prototype in the tumor institute of Hubei province, P. R. China from Jan. 2001 to Dec. 2001. Age range of patients was 18-76 (average age was 46). All primary NIR breast images were processed with the method of digital image processing and then analyzed. 76 cases estimated to be breast cancer were scheduled for biopsy. Final conclusion of diagnosis was based on the results of pathology.

2.2 Review of Principles of the Method

The adopted method employs a variety of digital image processing units, including Laplacian filters, rank value filters, edge detectors, gain units and summation units etc. The various units are combined to produce different operator, each operator produces a desired visual effect. Two of operators are applied to the NIR breast diagnosis.
In the following discussion, we assume that the image is a digital image. In an x-y coordinate system, a digital image is denoted by P.sub.(x, y), where x and y are the coordinates for the particular pixel in the image. P denotes the intensity of the pixel.

Laplacian Filter. An example of a Laplacian filter is given by equation (1):

$$P.sub.2\,(x, y)=4P.sub.1\,(x, y)-P.sub.1\,(x-1, y)-P.sub.1\,(x+1,y)-P.sub.1\,(x, y-1)-P.sub.1\,(x, y+1)\,. \quad (1)$$

The Laplacian filter overall has an effective image-sharpening or detail enhancement effect.

Rank Value Filter. All the pixels in the selected neighborhood are ranked from smallest to largest in intensity. The center pixel in the neighborhood is then replaced with the pixel value that has a specified rank. For example, a median rank filter replaces the center pixel with the pixel value that represents the middle or median rank.

Edge Detector. It outputs a high value when there is a sharp change in image intensity and outputs a low value in areas of constant intensity. The output of an edge detector is useful for emphasizing or de-emphasizing the edge content in an image.
Edge detectors in common use are the edge magnitude detector, the Sobel Edge Detector, the Compass Gradient Edge Detector, the Laplacian Edge Detector, the Roberts Edge Detector and the Difference of Gaussians Edge Detector.

Other Units. Gain unit is to simply multiply the intensity of each pixel by a constant gain, denoted G. This is presented by an equation (2):

$$P.sub.2\ (x, y)=GP.sub.1\ (x, y). \tag{2}$$

Summation Unit. It is to add two images together as, by the equation (3):

$$P.sub.3\ (x, y)=P.sub.1\ (x, y)+P.sub.2\ (x, y). \tag{3}$$

2.3 Appling the Method to NIR Breast Diagnosis

In some primary NIR breast images, malignant manifestation does not appear clearly and results in false negative diagnosis likely. On the contrary, some primary images of benign focus are similar to malignant lesion and results in false positive result likely. The above-mentioned digital image processing units have been combined to produce a variety of operators. Two of operators are applied to the NIR breast images to improve the reliability of breast diagnosis. One is known as glowing edges operator, which can produce a visual glowing edges effect and extract edges of the NIR breast image. Another is called as wrapping vessels operator, which can result in a visual effect of sharpening mammary blood vessels and highlight some tiny or fuzzy vessels. The pathologic patterns obtained by applying the two operators to NIR breast images express the structural and the vascular information of mammary tissue respectively.

The Glowing Edges Operator. Edges are elemental features of image. As to medical images, to extract structural features of the interest objects from images is required. An improved Sobel edge detector is the central unit of the glowing edges operator. In a digital image {P.sub. (x, y)}, Sobel edge detector is defined as equation (4):

$$\begin{aligned} A=&[P.sub.1\ (x-1, y+1) +2P.sub.1\ (x-1, y) +P.sub.1\ (x-1, y-1)]- \\ &[P.sub.1\ (x+1, y+1) +2P.sub.1\ (x+1, y) +P.sub.1\ (x+1, y-1)] \end{aligned} \tag{4}$$

$$\begin{aligned} B=&[P.sub.1\ (x-1, y-1) +2P.sub.1\ (x, y-1)+ P.sub.1\ (x+1, y-1)]- \\ &[P.sub.1\ (x-1, y+1)+2P.sub.1\ (x, y+1) +P.sub.1\ (x+1, y+1)] \end{aligned}$$

$$S(x,y)=(A^2+B^2)^{1/2}\ .$$

This method may lose some edges whose magnitude is less. To eliminate the possibility of data overflowing, we use an attenuation factor to divide the calculated result. By this means, we can get a non-anamorphic gray-level edge image and retain all edges of the image. This is presented by an equation (5):

$$S(x,y)=(A^2+B^2)^{1/2}\ /Scale\ . \tag{5}$$

The luminance contrast of the image is determined by the anatomic and pathologic state of the breast. The density of a normal breast is homogeneous and the images show homogeneous optical features. So the gray-value distribution of a normal NIR

breast image is ordered and the pattern formed by applying the glowing edges operator to a primary normal breast image can express regular edge lines. On the contrary, as to a malignant focus, no matter it is ductal carcinoma or lobular carcinoma, its structures has been greatly different from the other locations. So the gray-value distribution of a malignant tumor location is disordered and the processed image can express the pattern that extracted edge lines tend to be irregular compared with the other locations. A primary breast cancer image is shown in Fig.1A, in which there is a light shadow region (marked by an arrow; the denotation is adopted in all images). The processed image with the glowing edges operator is shown in Fig.1B. The corresponding region marked by the arrow shows the pattern that extracted edge lines tend to be disordered. In the other parts divided by the blood vessels, edge lines form well-regulated curves, some of which seem to be a series of concentric circles on the whole. These patterns show that the shadow part is a malignant tumor region while the others are normal.

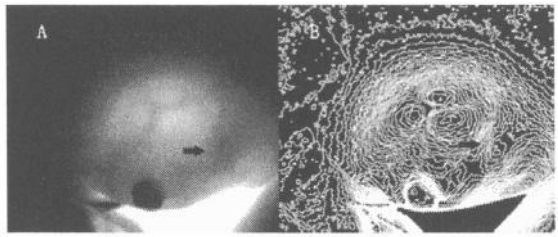


Fig. 1 Images of breast cancer. A is primary image, in which a light shadow region can be seen. B is the processed image with the glowing edges operator, in which the corresponding region shows the pattern that the edge lines tend to be disordered. In the other parts of B, edge lines form well-regulated curves, some parts seem to be a series of concentric circles on the whole.

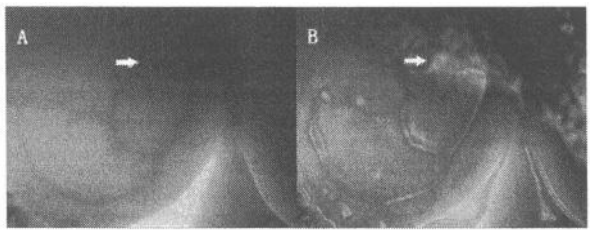


Fig. 2 Images of breast cancer. A is the primary image, in which there are some thick, crook blood vessels, but the tiny blood vessels are obscured by the shadow region. B is the processed image with the wrapping vessels operator, the tiny vessels and fuzzy vessels in the interest region is observed easily.

The Wrapping Vessels Operator. Malignant tumors might have high blood volume with low oxygen saturation since both a higher blood content and higher metabolism are necessary to achieve tumor growth in proliferating tumor tissue [7]. Increased blood volume associated with a rapidly growing tumor is the main cause of the increase of absorption in the malignant tumor relative to the background tissue [4]. In

the primary image of a malignant tumor, the blood vessels are generally thick, crooked and there are many tiny blood vessels obscured by the shadow region. The wrapping vessels operator can highlight some tiny vessels or fuzzy vessels in the shadow. Anisotropic gradient edge detector and median rank filter are the central units of the wrapping vessels operator. The primary image and the processed patterns of an early stage cancer are illustrated in Fig.2.

3 Results and Discussion

By means of the above-mentioned two operators, the pathologic patterns of 1466 cases were analyzed. 76 cases among them were estimated to be cancers, and 73 cases were certified to be malignant tumors by the pathological analysis of surgical sections. The other three cases were active hyperplasia, indistinct duct hyperplasia (precancerous lesions) and benign tumor respectively. Moreover, we diagnosed some benign neoplasms with the two operators, which accorded with the pathological results.

To sum up, in the course of diagnosing cases with the two operators, primary and processed images of different breast diseases expressed some specified features. Based on the analysis of these features, some false positive and false negative diagnoses were avoided. Representative cases presented in the following were sorted: malignant neoplasms and precancerous lesion or early stage cancer.

3.1 Malignant Neoplasm

Carcinoma of the breast is divided into noninvasive and invasive. Noninvasive carcinoma is epithelial proliferation confined to the terminal duct lobular units, has not invaded beyond the basement membrane and is therefore incapable of metastasis. Invasive breast carcinoma is breast tumor that has extended across the basement membrane and is thereby a lethal outcome. The cell abnormality of different types of carcinoma appears diverse characters, which determine the diversity of images. The malignant focus of a NIR breast image is based on features of the shadow, the blood vessels and the location relationship between shadow and blood vessels.

Firstly, features of the shadow: a typical primary image was shown as Fig.3A, the interest region was deep shadow; gray-value distribution of the shadow was not homogeneous, the center was deeper than the boundary and the boundary of the shadow was ambiguous generally. About 60% primary images were typical. In an untypical primary image shown as Fig.4A, the shadow was very faint, similar to the image of hyperplasia. In some cases, there only was a little starlike shadow owing to the smallness of the focus, it was shown as Fig.5A. The images processed by the glowing edges operator expressed that extracted edge lines of malignant focus tent to be irregular compared with the other locations of the image (shown as B of Fig.3-Fig.5).

Secondly, features of blood vessels as follows: 1) there were crooked, thick blood vessels in the periphery of shadow; 2) the shadow was surrounded by blood

vessels; 3) a section of blood vessel in the shadow appeared flocky; 4) there were tiny filiform blood vessels in the shadow; 5) blood vessels intersected or decussated in the shadow; 6) a thick blood vessel derived from inner part or edges of the shadow. The wrapping vessels operator can extract many tiny blood vessels obscured by the shadow region. They were shown in C of Fig. 3-Fig 5. In addition, the wrapping vessels operator made blood vessels in the periphery of shadow clearer than those in the primary image.

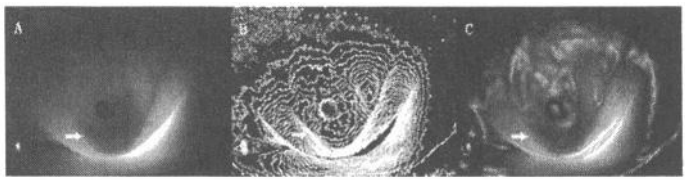


Fig. 3 Images of breast cancer. A is the primary image, in which the deep shadow near the nipple is surrounded by a thick blood vessel on one side. B is the processed image with the glowing edges operator, in which the edge lines in the corresponding region are scattered and unsystematic compared with the other parts. C is the processed image with the wrapping vessels operator, in which some separate blood vessels are extracted in the corresponding region.

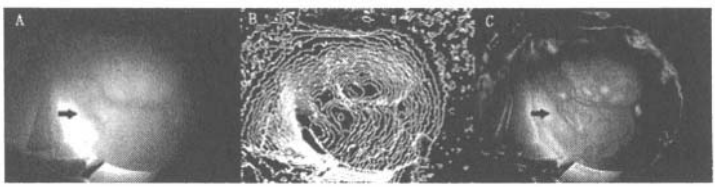


Fig. 4 Images of breast cancer. A is the primary image, in which blood vessels decussated in the light shadow. B is the processed image with the glowing edges operator, in which the edge lines in the corresponding region are a little disordered, compared with the other locations. C is the processed image with the wrapping vessels operator, in which extracted blood vessels in the corresponding region become clearer than those in A.

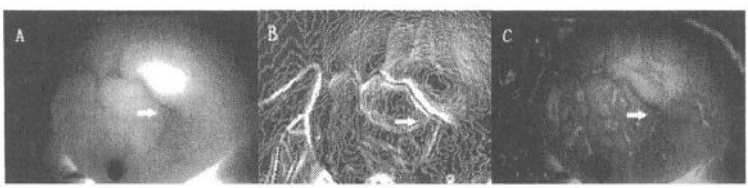


Fig. 5 Images of breast cancer. A is the primary image, in which a thick blood vessel derived from the light shadow. B is the processed image with the glowing edges operator, in which the edge lines in the corresponding region are disorganized compared with the other locations. C is the processed image with the wrapping vessels operator, in which some tiny separate blood vessels are extracted in the corresponding region.

3.2 Precancerous Lesion or Early Stage Cancer

Due to malignant focuses of some early stage cancer were too small to be shown, there is not typical malignant manifestation in primary images. The processed images with two operators expressed different typical features of pathologic changes in focuses.

Active Hyperplasia. There is a light shadow near the nipple in Fig.6A. In the processed image with the glowing edges operator (Fig.6 B), the edge lines in corresponding region are irregular. In Fig.6C, these blood vessels extracted in the interest region reveal some pathological changes. Therefore, the patient was estimated likely to suffer certain malignant changes. The pathological section acquired by the operation certified that it was active hyperplasia.

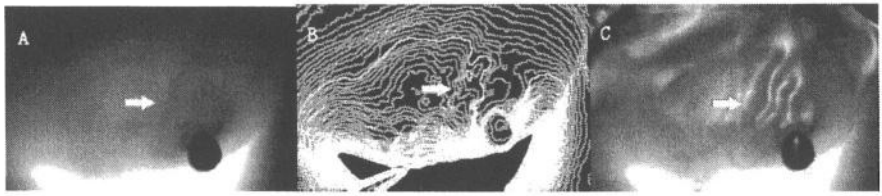


Fig. 6 Images of active hyperplasia. A is the primary image, in which the shadow near the nipple is very light. B is the processed image with the glowing edges operator, in which the edge lines in the corresponding region are distorted compared with the other locations. C is the processed image with the wrapping vessels operator, in which thick, crook blood vessels are extracted in the corresponding region.

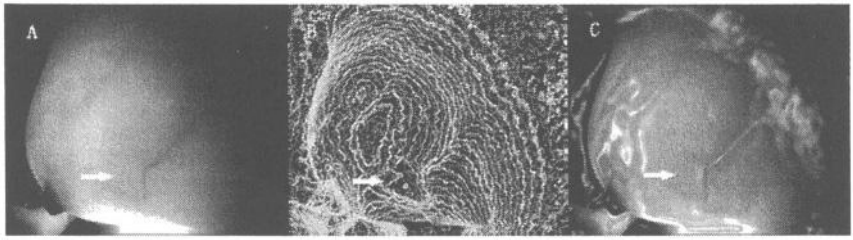


Fig. 7 Images of atypical duct hyperplasia. A is the primary image, in which there is a blood vessel passing through the light shadow region. B is the processed image with the glowing edges operator, in which the edge lines in the corresponding region are irregular compared with the other locations. C is the processed image with the wrapping vessels operator, in which blood vessels extracted in the interested region are unwonted.

Atypical Duct Hyperplasia. In the primary NIR image (Fig.7A), there is a blood vessel passing through the light shading and the borderline of the blood vessel is slightly coarse in the interest region. Fig. 10B shows the processed image with the glowing edges operator. The edge lines in corresponding region are slightly disordered and cannot form regular concentric circles partly. The feature in Fig.7B demonstrated that tissues in the interest region are abnormal. Fig.7C shows the processed

image with the wrapping vessels operator. The blood vessels extracted in the interested region are unwonted. According to the characters of the three images, we estimated that there are some pathological changes in the breast and the changes are likely to be malignant. The pathological section of the breast acquired by the operation certifies that it is an indistinct duct hyperplasia, a precancerous lesion.

4 Conclusions

We have diagnosed 1466 clinical patients with NIR breast imaging. Applying the method of digital image processing to these NIR breast images, we analyzed the primary and processed images by comparison. We estimated 76 cases to be cancers among 1466 clinical cases, and 73 cases were certified to be malignant tumors by the pathological examination. Without the method, only based on primary images, the possibility of false positive and false negative was high. Bases on the method, we have avoided the performance of some biopsies for benign lesions and founded some precancerous lesion or early stage cancer. It can be concluded that with the method of digital image processing, the reliability of diagnosing mammary diseases can be raised obviously.

References

1. M. Cutler.: Transillumination as An Aid in The Diagnosis of Breast Lesion. Surg Gyneco Obstet. 48 (1929) 721-730
2. Ohlsson B, Gundersen J, Nilsson DM.: Diaphanography: A Method for Evaluation of The Female Breast. World J Surg. 4 (1980) 701-706
3. Watmough DJ.: Transillumination of Breast Tissues: Factors Governing Optimal Imaging of Lesions. Radiology. 174 (1983) 89-92
4. Jiang H, Iftimia NV, Xu Y, Eggert JA, Fajardo LL, Klove KL.: 2002.Near-infrared Optical Imaging of The Breast With Model-based Reconstruction. Acad Radiol. 9 (1995) 186-194
5. J. R. Lakowicz and K. Berndt: Frequency Domain Measurement of Photon Migration in Tissue. Chem. Phys. Lett., 16 (1990) 246-252
6. B. Chance, et al.: Comparison of Time Resolved and Unresolved Measurements of Deoxyhaemoglobin in Brain. Proc. Natn. Acad. Sci. USA, 8 (1988) 4971-4975
7. Thomsen S, Tatman D.: Physiological and Pathological Factors of Human Breast Disease That Can Influence Optical Diagnosis. Ann N Y Acad Sci. 838 (1998) 171–193

Comparison of Phase-Encoded and Sensitivity-Encoded Spectroscopic Imaging*

Min Huang, Songtao Lu, Jiarui Lin, Yingjian Zhan

MRSI Research Group, Institute of Biomedical Engineering,
Huazhong University of Science and Technology, Wuhan, 430074, China
hmrose@163.net lstsam@yofc.com
Linjiaru@hust.edu.cn zhanyingjian@hotmail.com

Abstract. MR Spectroscopic imaging plays a more and more important role in clinical application. In this paper, we make comparison on two MRSI technologies and data reconstruction methods. For the conventional phase-encoded spectroscopic imaging, the data reconstruction using FFT is simple, but the data acquisition is very time consuming and thus prohibitive in clinical settings. Sensitivity-encoded SI is a new parallel approach of reducing the acquisition time by reducing the necessary spatial encoding steps with multiple coils. It uses the distinct spatial sensitivities of the individual coil elements to recover the missing encoding information in reconstruction. Fourfold reduction in scan time can be achieved when the factor of R_x and R_y are both 2, with no compromise in spectra and spatial resolution. These improvements in data acquisition and image reconstruction provide a potential value of metabolic imaging using SENSE-SI as a clinical tool.

1 Introduction

MR spectroscopic imaging (MRSI, SI) is a completely noninvasive imaging method. In contrast to magnetic resonance imaging (MRI), MRSI can present information in the form of metabolite maps, which represent not only simply anatomy but also local metabolic states or local tissue abnormalities [1]. It shows great promise for use in basic physiological research and for clinical imaging of metabolic function. SI has been proposed as a method to localize and assess brain tumors [2], multiple sclerosis, and temporal lobe epilepsy [3].

In vivo SI suffers generally from long imaging time and poor signal-to-noise ratio (SNR), because of a combination of a weak MR signal and low metabolite concentrations. Low SNR limits the technique's ability to detect metabolite abnormalities in subjects. Therefore, an increase of speed and SNR is a key factor in the success of many MRSI applications.

* Grant sponsor: National Natural Science Foundation of China, Grant number: 30070226. University Foundation of HUST, Grant number: 0101170054.

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 70-77, 2004.

For the conventional phase encoding spectroscopic imaging (PHASE-SI), the data reconstruction using FFT is simple. But the data acquisition is very time consuming and thus prohibitive in clinical settings [4]. In the past ten years, many approaches have been proposed and implemented for faster MRSI. Correspondingly, a lot of reconstruction methods have also been proposed. The majority of fast SI methods are based on fast MRI sequences to accelerate data sampling, in which the signal is acquired under a time-varying readout gradient. All signals from at least one whole slice of *k*-space are acquired within the TR, e.g., from a $(k_x - t)$ plane using echo planar spectroscopic imaging (EPSI) [5] or from a $(k_x - k_y)$ plane using spiral SI [6], and so on. Because sampled data points are not on a rectilinear grid, special reconstruction algorithms are required. EPSI reconstruction uses shift of odd and even echoes [5]. Spiral SI uses gridding FFT [7]. However, there are some restricts. EPSI requires rapidly switching gradient field and strong power provide system. For spiral method, the smoothing effects of gridding kernel will deconvolve in the image domain, thus will affect the resolution of image. And now, a new parallel imaging technique based on sensitivity-encoded (SENSE) has arisen [8-11]. SENSE-SI applies coil arrays for parallel data acquisition to reduce the acquisition time by reducing the number of sampled *k*-space points in the spatial frequency domain [12]. In the process of data reconstruction, it uses the distinct spatial sensitivities of the individual coil elements to recover the missing encoding information [13].

2 Theory

The collected raw signal equation in SI is given by

$$M(k_x, k_y, t) = \sum_{k=1}^{K} \int_x \int_y m_k(x, y) \exp[i2\pi(k_x x + k_y y)] \exp[(-\lambda_k + i\omega_k)t] dx dy \tag{1}$$

where $m_k(x, y)$ is the density in the position (x,y,z) and λ_k is the decay constant of the *k–th* metabolite component, $\lambda_k = 1/T_{2k}$. ω_k is the chemical shift frequency, $\omega_k = -2\pi\gamma B_0 \delta_k = 2\pi f_k$. $M(k_x, k_y, t)$ is the acquisition signal, i.e. the raw data.

We can see that the raw signal is possible by appropriately sampling the *k*-space and time domain (k_x, k_y, t). *K-space* (k_x, k_y) is the Fourier domain of the space (x, y), i.e. the data-sampling trajectory. For different SI technologies, there are various forms of *k*-space sampling schemes, which can be realized by gradient design.

$$k_x(t) = \gamma \int_0^t G_x(\xi) d\xi \tag{2}$$

$$k_y(t) = \gamma \int_0^t G_y(\xi) d\xi \tag{3}$$

Not only are the spatial distributions of the spins important but also are the spectral components with each spatial position. Spectral components can be gathered by collecting the time direction. The ultimate spatial distribution of spectral information of

the imaged object get is $m(x,y,f)$, which can be acquired by mathematical algorithm to the raw data $M(k_x,k_y,t)$.

2.1 PHASE-SI

In 2D conventional PHASE-SI, the FID signals are collected along t axis at a fixed point (k_x,k_y) in the absence of a gradient. Therefore the spectral and spatial information is completely separated. The spatial information, which has a regular distribution in spatial-frequency (or k-space) domain, is obtained by gradient phase encoding in two spatial dimensions. The spectral information is sampled directly in an additional t dimension [4].

To reconstruct the data, a 3D Fourier transform applied on the (k_x,k_y,t) domain is sufficient to resolve a spectrum in each pixel in a 2D image. First, we apply apodization and 1D Fourier transform for the k-space FID signal in eq.1 to get the k-space spectra information:

$$m(k_x,k_y,f)=\sum_{t=0}^{N-1}\exp(-f_A t)S(k_x,k_y,t)\exp(i2\pi ft/N) \tag{4}$$

where, $f = -N/2,...,0,...,N/2-1$, f_A is apodization frequency. Second, we apply 2D Fourier transform in spatial dimensions, and finally get the space spectra information:

$$m(x,y,f)=\sum_{1}^{N_{kx}}\sum_{1}^{N_{ky}} m(k_x,k_y,f)\exp[i2\pi(\frac{k_x.x}{N_{kx}}+\frac{k_y.y}{N_{ky}})] \tag{5}$$

The product of the repetition time (TR) and the phase encoding steps determines the scan time. For 2D 32*32 spatial resolution with TR 2000ms, the total scan time is about 34 minutes. So the phase encoding SI cannot be suitable for clinical examination because of the long acquisition tune.

2.2 SENSE-SI

SENSE-SI uses a new approach based on SENSE-MRI technology [8], which is different from other fast imaging technologies. It applies coil arrays for parallel data acquisition to reduce the acquisition time and permit decreasing the sampling density in k-space. For most efficient scan time reduction in conventional SI, SENSE can be applied to both spatial phase encoding dimensions, x and y. The FOV is reduced by a factor R_x in x direction and a factor R_y in y direction in SENSE-SI. Only a fraction of k-space positions is sampled as compared with full k-space sampling, leading to scan time reduction by the factor $R = R_x * R_y$. Thus, if the full FOV is to be resolved by $n\times n$ spectra, only $n/R_x \times n/R_y$ individual signals need to be sampled. In the imaging domain, this sampling scheme corresponds to a $n/R_x \times n/R_y$ grid within a reduced FOV. Because SENSE increases the distance between each line in k-space by reduc-

ing the number of phase encoding steps, the negative effect of under-sampling is that the image will be aliased in the phase encoding direction after Fourier reconstruction [10]. To create full unaliased FOV metabolite maps, signal contributions of different coils must be separated in SENSE reconstruction. SENSE reconstruction exploits the fact that each signal contribution is weighted according to the local sensitivity of each coil. After discrete Fourier transform in all three dimensions, an aliased image needs to be unfolded for every sampling frequency in the spectral direction. The metabolic image without aliased artifacts will be achieved after unfolding the raw images with superposition of pixels.

Let n_c be the number of coils, n_p be the number of pixels superimposed on each pixel in the reduced FOV image. Then, for each pixel in the reduced FOV image, a $n_c * n_p$ sensitivity matrix S is created.

$$S(r_\lambda) = \begin{pmatrix} s_1^1 & s_1^2 & \ldots s_1^{n_p} \\ s_2^1 & s_2^2 & \ldots s_2^{n_p} \\ \vdots & \vdots & \ldots \vdots \\ s_{n_c}^1 & s_{n_c}^2 & \ldots s_{n_c}^{n_p} \end{pmatrix} \tag{6}$$

where, $S_{n_c}^{n_p}$ is the coil sensitivity value in the superimposed pixel n_p form coil n_c. r_λ denotes the position that pixel λ has in the reduced FOV.

Let α be a vector containing the image values that the chosen pixel λ has in the n_c coil images. The vector v contains the unfolded pixel values for the original superimposed positions in the full-FOV. The relation between v, α and S is:

$$\begin{pmatrix} s_1^1 & s_1^2 & \ldots & s_1^{n_p} \\ s_2^1 & s_2^2 & \ldots & s_2^{n_p} \\ \vdots & \vdots & \ldots & \vdots \\ s_{n_c}^1 & s_{n_c}^2 & \ldots & s_{n_c}^{n_p} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_{n_p} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{n_p} \end{pmatrix} \Rightarrow Sv = \alpha \tag{7}$$

So, for each spectral sampling point λ, signal separation can be achieved by

$$V_\lambda = U\alpha_\lambda \tag{8}$$

Then the reconstruction unfolding matrix U is given by

$$U = (S^H \psi^{-1} S)^{-1} S^H \psi^{-1} \tag{9}$$

where H denotes the transposed complex conjugate and Ψ is there receiver noise matrix. The matrix Ψ is determined experimentally, using

$$\psi_{\alpha\alpha'} = \overline{\eta_\alpha \eta_{\alpha'}}^* \tag{10}$$

Where η_α denotes the pure noise output of the $\alpha - th$ receiver channel, the bar denotes time averaging, and the asterisk * denotes the complex conjugate.

3 Materials and Methods

Data Acquisition of phantom and in vivo
We apply two kinds of raw data of PHASE-SI and SENSE-SI to compare the two imaging technologies and image reconstruction methods. The experiments were conduced on a 1.5T Philips Intera whole body scanner (Gyroscan ACS-NT) from the institute of biomedical engineering, University of ETH Zurich, Switzerland. The experiments use a receiver array of six surface coils, two of which are circular and four are rectangular. The coil array is arranged around the phantom or the patient's head. The phantom are full of CRE (10mmol/l), around which are three glass spheres filled with LAC (10mmol/l), LAC and NAA (5mmol/l) and NAA (10mmol/l) respectively (Fig3.a). The field of view (FOV) is 230mm, which is divided into 32*32 voxels with a slice thickness of 1.5cm. The nominal voxel volume is 0.77cc. The FOV of in vivo volunteer is 220mm, which is divided into 24*24 voxels with a slice thickness of 2cm. The nominal voxel volume is 1.7cc. 256 samples are acquired in per spectrum over a bandwidth of 1000Hz, using a TR of 1500 ms for phantom (1700 ms for in vivo), a TE of 288 ms. The spectral resolution is 4 Hz. 2D PHASE-SI and SENSE-SI measurements are collected in phantom and in vivo, with a reduction factor (R_x or R_y) of two in both spatial dimensions for SENSE.

Data preprocessing
Cosine filtering - For the raw data with the resolution of 32*32*256 for phantom and of 24*24*256 for in vivo in PHASE-SI, and of 16*16*256 for in phantom and of 12*12*256 for in vivo in SENSE-SI, cosine filtering is first applied to two spatial dimensions (k_x, k_y) in k-space for each coil before FT reconstruction in three dimensions. The cosine filter is also referred as spatial apodization, which is used to reduce the ringing artifact. Gaussian filtering - Gaussian (exponential) function is multiplied by each time-domain FID or echo signal for apodization in the frequency domain. It can reduce the noise components and improve the SNR.

Data reconstruction
After preprocessing, the data of PHASE-SI can be reconstructed using Fast Fourier transform (FFT) directly in three dimensions to get the spectrum. For SENSE-SI, extra SENSE reconstruction is needed to separate six-coil signal contributions to get the unaliased spectrum in full FOV after FFT.

Postprocessing
After Reconstruction, the spectrum must be corrected with B0-map. Zero order phase correction and polynomial baseline correction are also used in the spectrum.

Metabolite image and data interpolation
After getting the spectrum, metabolic images of NAA, CHO, CRE and LAC are obtained by integration of the modulus peak respectively. The resultant metabolic images are Fourier interpolated to 256*256 pixels with the same resolution as MRI.

4 Result

We have designed and implemented software of reconstruction algorithm for comparison of PHASE-SI and SENSE-SI using IDL (Interactive Data Language) version 6.0. Fig.1 is the interface of the SI reconstruction software. It includes data preprocessing, data reconstruction and postprocessing for the data of two acquisition styles. The spectrum in different voxel and the image of four metabolites (including NAA, CHO, CRE and LAC) can be acquired after the corresponding processing.

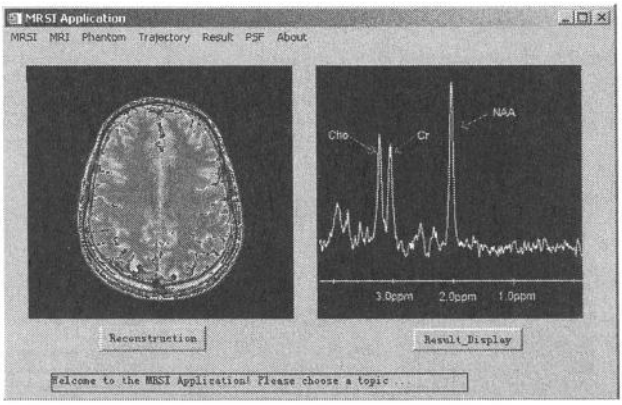


Fig. 1. The interface of reconstruction software using IDL

For in phantom study, scan time for PHASE-SI is about 22 minutes and 33 minutes, which for SENSE-SI is 5 minutes and 37 seconds.

Fig.2 shows the reconstruction results of the phantom using PHASE-SI and SENSE-SI respectively. Fig.2a shows the scout image with the resolution of 256*256. The reconstructed spectrum (Fig.2b) in the voxel of the flag in the middle glass sphere (Fig.2a) shows that there are metabolites of NAA and LAC with the same concentration, but there is signal loss in the spectra of SENSE-SI. Fig.2c shows the distribution of NAA, we can see that there are solutions of NAA only in the middle and the right glass sphere. The metabolite image and spectra (right) of SENSE-SI have no compromise in the resolution of spectra compared with that (left) of PHASE-SI, except the lower SNR with reduced scan time.

For in vivo study, scan time for PHASE-SI is about 14 minutes and 2 seconds, which for SENSE-SI is 3 minutes and 37 seconds. Fig.3 shows the reconstruction results of the patient with a grade II astrocytoma using PHASE-SI and SENSE-SI respectively. Fig.3a shows the anatomic image of the patient with the resolution of 256*256. Fig.3b and Fig.3c shows the distribution of LAC, there is metabolite of LAC only in active tumor tissue, which shows different gray value in the left position of tumor tissue in the scout image. The characteristic metabolic pattern of the astrocytoma is clearly visible with a fourfold scan time reduction using SENSE-SI (right), as well as PHASE-SI (left). But the reduced scan time and the broadening of the Spatial Response Function (SRF) led to visibly reduced SNR in the SENSE-SI images.

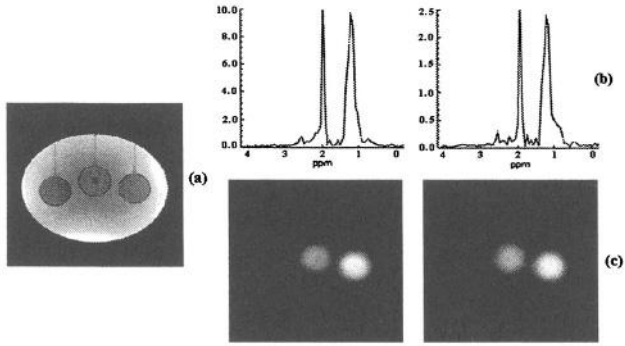


Fig. 2. The reconstruction result for phantom: (a) the scout image of phantom with the resolution of 256*256, (b) the reconstructed spectrum signal in flag with PHASE-SI (Left) and SENSE-SI (Right), (c) the metabolite image of NAA with PHASE-SI (Left) and SENSE-SI (Right).

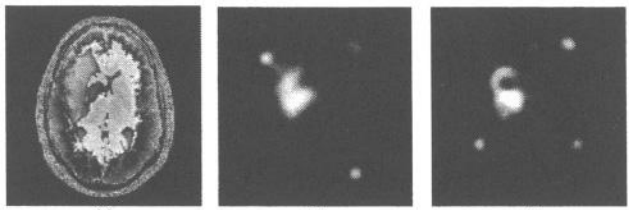


Fig. 3. The reconstructed result for in vivo: (a) the anatomic image of a patient with grade II astrocytoma, (b) and (c) shows the metabolite image of LAC with PHASE-SI and SENSE-SI respectively.

We can see that the scan time for SENSE-SI can be reduced by 4 compared with PHASE-SI when the factor of R_x and that of R_y are both 2. After SENSE reconstruction, the aliasing artifacts are eliminated. The reconstructed results for SENSE-SI are in good agreement with that for PHASE-SI, except that the lower SNR in the former.

5 Discussion and Conclusion

The main difference between the PHASE-SI and SENSE-SI technologies is the FOV for acquisition. PHASE-SI samples the full information in the two spatial dimensions (k_x, k_y). The data reconstruction using FFT is simple, but the data acquisition is very time consuming and thus prohibitive in clinical settings. In SENSE-SI, sensitivity encoding can be applied to both spatial phase encoding dimensions, *x* and *y,* and only a fraction of k-space positions is sampled as compared with full *k*-space sampling, leading to scan time reduction by the factor $R = R_x \times R_y$, SENSE-SI reconstruction uses the distinct spatial sensitivities of the individual coil elements to recover the missing encoding information and get the unaliased FOV metabolite maps. There is no loss of

spectral resolution in SENSE-SI compared with many of the other fast SI methods, and no restrictions in the spectral bandwidth or echo time. And SENSE-SI reduce the acquisition time by using multiple coils rather than a specific pulse sequence, so it can be also applied into the other fast SI to improves the speed highly. In conclusion, SENSE-SI improves the data acquisition speed, and the new data reconstruction also makes the image quality qualified to be view. Both of them provide a potential value of metabolic imaging as a clinical tool.

Acknowledgments. The authors are grateful to Dr. Ulrik Dynak from the institute of biomedical engineering, University of ETH Zurich, Switzerland, for providing the *in vivo* Phase encoding SI and SENSE-SI raw data used in this study and for helpful discussions.

References

1. Kirn T.F.: Magnetic Resonance Spectroscopy May Hold Promise in Studying Metabolites Tissues. J. Amer. Med. Asso. 261 (1989) 1102-1103
2. Luyten P.R., Marien A.J., Heindel W., Gerwen P.H.: Metabolic Imaging of Patients with Intracranial Tumors: H-1 MR Spectroscopic Imaging and PET. Radiology. 176 (1990) 791–799
3. Arnold D.L., Matthews P.M., Francis G.S:Proton Magnetic Resonance Spectroscopic Imaging for Metabolic Characterization of Demyelinating Plaques. Ann. Neurol. 31 (1992) 235–241
4. Brown T.R., Kincaid B.M., Ugurbil K.: NMR Chemical Shift Imaging in Three Dimensions. Proceedings National Academy of Sciences USA, 79(1982) 3523~3526
5. Posse S., Tedeschi G., Risinger R., Ogg R., Bihan D.: High Speed 1H Spectroscopic Imaging in Human Brain by Echo Planar Spatial-spectral Encoding. Magn. Reson. Med. 33 (1995) 34-40
6. Adalsteinsson, E., Irarrázaval, P., Topp, S., Meyer, C.H.: Volumetric Spectroscopic Imaging with Spiral-based k-space Trajectories. Magn. Reson. Med. 39 (1998) 889-898
7. Rasche V., Proksa R., Sinkus R., Bornert P.: Resampling of Data Between Arbitrary Grids Using Convolution Interpolation, IEEE Transactions on Medical Imaging, 18 (1999) 385-392
8. Pruessmann K.P., M.Weiger, MB. Scheidegger, P. Boesiger. SENSE: Sensitivity Encoding for Fast MRI, Magn Reson Med 42 (1999) 952-962
9. Weiger M., Pruessmann K.P., Boesiger P.: 2D SENSE for Faster 3D MRI. Magn Reson Mater Phy 14(2002)10-19
10. Dydak U., Weiger M., Pruessmann K.P., Meirer D., Boesiger P.: Sensitivity Encoded Spectroscopic Imaging. Magn Reson Med 46 (2001) 713-722
11. Dydak U., Pruessmann K.P., Weiger M, Tsao J., Meier D., and Boesiger P.: Parallel Spectroscopic Imaging With Spin-Echo Trains. Magn. Reson. Med. 50 (2003) 196–200
12. Pruessmann K.P., Weiger M., Bornert P., and Boesiger P.: Advances in Sensitivity Encoding with Arbitrary k-space Trajectories. Magn. Reson. Med. 46 (2001) 638–651
13. Kyriakos W.E., Panych L.P., Kacher D. F., Westin C.: Sensitivity Profiles From an Array of Coils for Encoding and Reconstruction in Parallel (SPACE RIP), Magn. Reson. Med.44 (2000) 301–308

Detection and Restoration of a Tampered Medical Image

Jun-Chou Chuang and Chin-Chen Chang

Department of Computer Science and Information Engineering,
National Chung Cheng University,
160, San-Hsing, Min-Hsiung, Chaiyi 621, Taiwan.
E-mail:ccc@cs.ccu.edu.tw
Tel: 886-5-2720411 ext. 33100
Fax: 886-5-2720859

Abstract. This paper presents a recoverable image tamper proofing technique using the symmetric key cryptosystem and vector quantization for detecting and restoring of a tampered medical image. Our scheme applies the one-way hashing function and the symmetric key cryptosystem to the host image to generate the verification data. To recover the tampered places, the host image is compressed by vector quantization to generate the recovery data. Once an intruder has modified the host image, our scheme can detect and recover the tampered places according to the embedded verification data and the recovery data. Besides, the proposed scheme can withstand the counterfeit attack.
Keywords: Authentication, medical imaging, watermarking, confidentiality

1 Introduction

The Internet provides a fast and convenient way for electronic patient records (EPRs) to be transmitted or received by hospitals and clinics. Generally, EPR contains the diagnostic report, prescriptions, histological, and diagnostic images. Because each EPR is a very private material for a patient, EPR transmitted through the Internet must be protected by some security mechanisms.

Image tamper proofing, also called *image authentication,* is a kind of image protection mechanisms. The goal of image tamper proofing is to identify the integrity of the digital images. Image authentication can be classified into hard authentication and soft authentication. The main difference between hard and soft authentications is that soft authentication allows modifications, but hard authentication does not. Two common approaches for image authentication are the digital signature approach [3,6,9] and the watermarking approach [2,7,8]. The digital signature approach is suitable for hard authentication because the authentication codes are saved as an independent file. The watermarking approach is suitable for soft authentication scheme due to that the watermarking approach embeds the authentication codes into the image directly.

In this paper, we propose a watermarking-based image tamper proofing scheme for medical image authentication. Our scheme divides an host image into blocks of equal size. The recovery data and the verification data are generated from each block by using vector quantization [4] and the symmetric key cryptosystem [5], respectively. The verification data and the recovery data are then embedded into the least significant bits of each pixel of the corresponding block.

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 78–85, 2004.

The remainder part of this paper is organized as follows. In Section 2, we review the vector quantization and the counterfeit attack. In Section 3, our proposed scheme is presented. The experiments and security analyzes are shown in Section 4. Finally, the conclusions are given in Section 5.

2 Previous Work

Here, we will review the vector quantization (VQ) compression and the counterfeit attack [1] on the block-based watermarking scheme in Subsections 2.1 and 2.2, respectively.

2.1 Vector Quantization

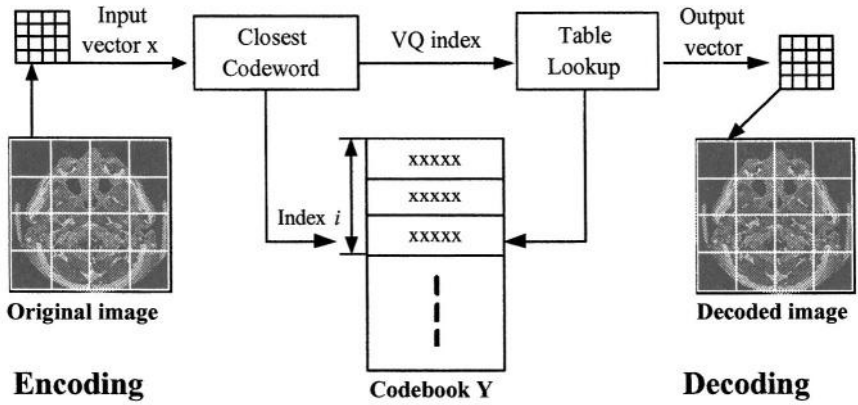


Fig. 1. The block diagram of VQ encoding/decoding

VQ is a very popular gray-level image compression technique because of its low bit rate and simple encoding/decoding. The block diagram of VQ encoding/decoding is shown in Fig. 1. In the encoding phase, the codebook is searched to find the closest codeword of the input vector x, and then the index pointing to the closest codeword is transmitted to the decoder. The codebook is generated by the Linde-Buzo-Gray (LBG) algorithm [4], To find the closest codeword for each image vector, the squared Euclidean distance is applied, and it is defined as follows.

$$d(x, y_i) = \sum_{j=1}^{k} (x_j - y_{ij})^2, \tag{1}$$

where x_j and y_{ij} denote the j-th elements of x_j and the j-th elements of y_i in codebook Y, respectively. If codeword y_t the closest codeword for the image vector x then

$d(x, y_t)$ is the smallest among $d(x, y_i)$'s for $i = 1, 2, ..., N_c$ and $i \neq t$. Here N_c is the codebook size. Then the encoder sends the VQ index y_t to the decoder. In the decoding phase, the decoder only needs to do the table look-up procedure to reconstruct the encoded image.

2.2 Block-Based Counterfeit Attack

Holliman and Memon [1] proposed the counterfeit attack on block-based independent watermarking scheme. The term "block-based independence" means the generation of a watermarked block does not refer to another image block. A watermarked block generation only depends on its original block, a watermark block, and an insert key $\bar{K}$. If the watermarking scheme is based on block-based hiding, then its watermarked image could be counterfeited.

The same watermark can be extracted from different watermarked blocks with the same key $\bar{K}$. Those watermarked blocks with the same watermark are called $\bar{K}$-*equivalent*. Attackers can collect a set of watermarked blocks from many watermarked images to generate a codebook $C = \{C_1, C_2, ..., C_\lambda\}$ according to $\bar{K}$-*equivalence*, where λ is the number of different possible watermark patterns.

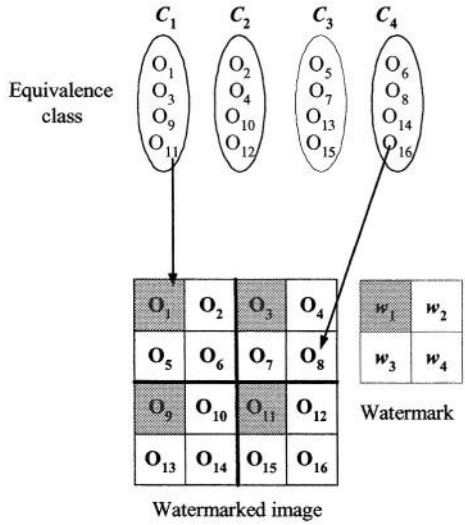


Fig. 2. An example of counterfeit attack

An example of counterfeit attack is shown in Fig. 2. Suppose we want to counterfeit two watermarked blocks o_3 and o_6, where $o_3 \in C_1$ and $o_6 \in C_4$. The attacker can choose $o_1 \in C_1$ and $o_8 \in C_4$ to replace o_3 and o_6 because they are belong to the same equivalence class. Hence, the attacker can create two forged blocks by the equivalence class. Even though two watermarked blocks o_3 and o_6 have been modified, the verification procedure still cannot detect the modified blocks.

3 The Proposed Scheme

The goal of our scheme is to provide a simple and secure image tamper proofing scheme which can detect and recovery the tampered places. Our scheme consists of two procedures, the signing procedure and the verification procedure. The details are described as follows.

3.1 The Signing Procedure

Given a gray-level medical image O, where $O = \{O_{ij} | 1 \leq i \leq M, 1 \leq j \leq N, O_{ij} \in \{0, 2^r - 1\}\}$ and r is the number of bits used to represent a pixel. The original image O is first divided into non-overlapping blocks of 16 × 16 pixels, where $O = \{o_1, o_2, ..., o_\tau\}$ and $\tau = \frac{M \times N}{16 \times 16}$. Next, the two least significant bits LSB_1 and LSB_2 of each gray pixel are set to be zero. Thereupon, each block o_i in O is converted to $\bar{o}_i$. In our scheme, the recovery data and the verification are embedded into the LSB_1 and LSB_2 of each gray pixel, and shown in Fig. 3.

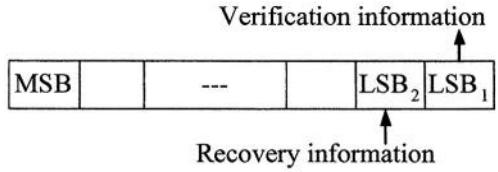


Fig. 3. Recovery and verification information of a pixel

The first step of the signing procedure is to generate the recovery data. For each input block $\bar{o}_i$, a block is divided into 16 sub-blocks of 4 × 4 pixels and then we use VQ to encode each sub-block. In order to obtain a better image quality of the decoded image, only one image O is selected to train a codebook. The codebook can be saved as an independent file or it can be placed in the header of the image file. Since 16 sub-blocks are encoded by VQ, the total number of recovery bits of a block is $(16 \times log_2 Nc)$, where Nc is the codebook size. The recovery bits are then sequentially embedded into the LSB_2 of each pixel of a block. In order to prevent the damage of the recovery bits of a block, the recovery bits are not embedded into the same block. They are randomly embedded into another block by using a pseudo-random number generator (PRNG) with a secret seed SD_1.Each block $\bar{o}_i$ is then converted to $\hat{o}_i$, where $i = 1, 2, ..., \tau$.

After embedding the recovery data into the host image, the second step is to generate the verification data. For each input block $\bar{o}_i$, the one-way hash function, e.g. MD5, is employed to generate a 256-bit hash value. The hash formula is as follows.

$$h_i = Hash(\hat{o}_i || P || T || k_i). \tag{2}$$

Here P is the patient identification, T is the doctor's seal, and k_i is a random number which is generated by a PRNG with a secret seed SD_2. The random number k_i is

used to prevent the VQ counterfeit attack. Then the verification data s_i of a block $\hat{o}_i$ is generated by computing the following formula.

$$s_i = Encrypt(h_i, SK). \tag{3}$$

where SK is the secret key of DES [5]. DES encryption does not cause the data expansion. The verification data s_i of 256 bits are hidden into the $\mathbf{LSB}_1$ of block $\hat{o}_i$ itself. Finally, a signed image is generated. An example of our signing procedure is shown in Figs. 4 and 5, and assume that the block encoding size is 2×2.

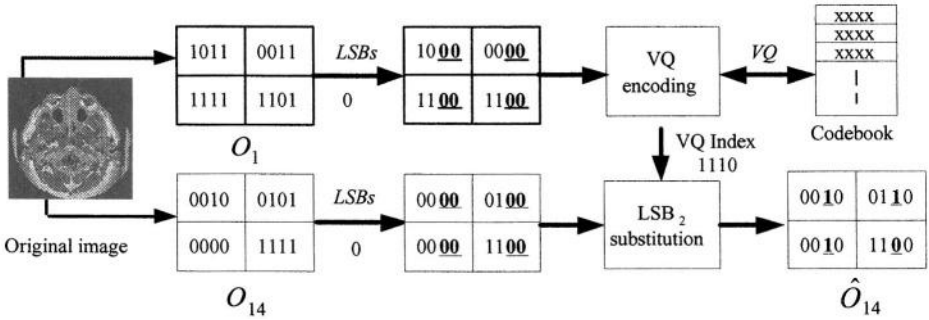


Fig. 4. Generation of the recovery data

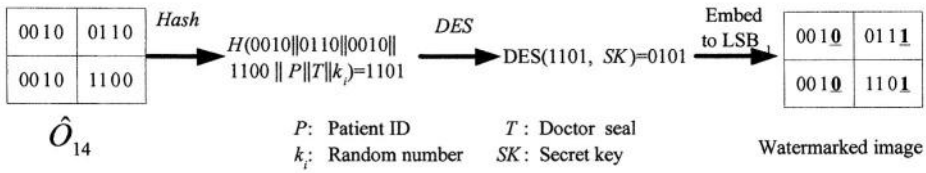


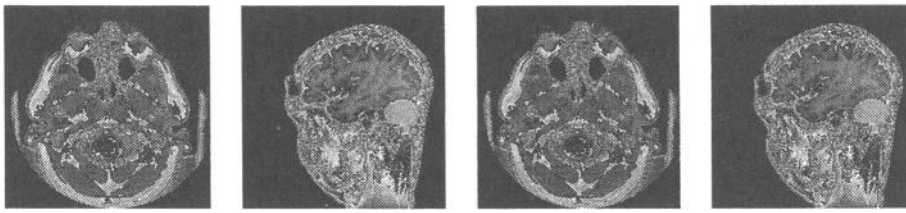
Fig. 5. Generation of the verification data

3.2 The Verification Procedure

The verification procedure is to check whether a test image O' is modified or not. When the detection scheme confirms an image block has been tampered with, the recovery process is then performed to recovery the modified block. In order to verify the integrity of each block of a test image, we need to execute the signing procedure to generate the new verification data of each block. For each block, the new verification data are compared with the embedded verification data of the same block. If the new verification data and the embedded ones are not identical, it denotes that this block has been modified.

4 Experimental Results and Analyzes

Two test gray-level medical images "CT" and "MRI" of 512 × 512 pixels and 256 gray levels (8 bits/pixel), as shown in Figs. 6(a)-(b), are used in our experiments. The codebook design is based on the LBG algorithm [4] with 1024 code vectors (N=1024), and the VQ encoding size is set to 4 × 4 pixels. Besides, size of a modified region in our scheme is a block of 16 × 16 pixels.



(a) CT (b) MRI (c) PSNR=45.44 dB (d) PSNR=45.01 dB

Fig. 6. (a)-(b) Original images, and (c)-(d) watermarked images

The human eyes and the peak-signal to noise rate (PSNR) are two approaches to estimate the image quality. The PSNR formula is defined as follows:

$$PSNR = 10 \times \log_{10} \frac{(2^r - 1)^2}{MSE} dB, \tag{4}$$

where r is the number of bits used to represent a pixel. The mean square error (MSE) of an $N \times N$ gray-level image is as follows.

$$MSE = \frac{1}{N \times N} \sum_{i=1}^{n} \sum_{j=1}^{n} (x_{ij} - \bar{x}_{ij})^2. \tag{5}$$

Here, X_{ij} denotes an original pixel value, and $\bar{x}_{ij}$ denotes the processed pixel value.

In the detection phase, a block is called a tampered block if the verification data of this block are not identical to the original ones. Thus, our scheme can not allow any kind of modifications such as malicious manipulation or involuntary manipulation. The involuntary manipulation includes JPEG compression, smoothing, sharping, and etc. The medical images are often compressed by the lossless image compression in order to save the storage space with a good image quality, Therefore, our scheme is very suitable for lossless medical image authentication.

In the recovery phase, if the detection scheme confirms an image block has been tampered with, the recovery work is then performed. In order to prevent the damage of the recovery data of a block, the recovery data are not embedded into the same block. They are randomly embedded into another block by using a pseudo-random number

generator with a secret seed. Our system can successfully recover a tampered block if the retrieved recovery data are correct, which denotes the backup block is not modified.

The watermarked images "CT" and "MRI" are shown in Figs. 6(c)-(d). The PSNR values of these two watermarked images are 45.44 dB and 45.01 dB, respectively. Two tampered images are shown in Figs. 7(b) and 7(f). The detected results are shown in Figures 7(c) and 7(g). The recovered results are shown in Figs. 7(d) and 7(h). From Figs. 7(d) and 7(h), it is obvious that the recovered image by using VQ is very similar to the host image. Since the image quality of the recovered image in our scheme is affected by the codebook size. If a smaller codebook size is used, the image quality of the watermarked image will be improved, but the image quality of the recovered image will be deteriorated. The codebook size appropriate to the experiments is 512 or 1024.

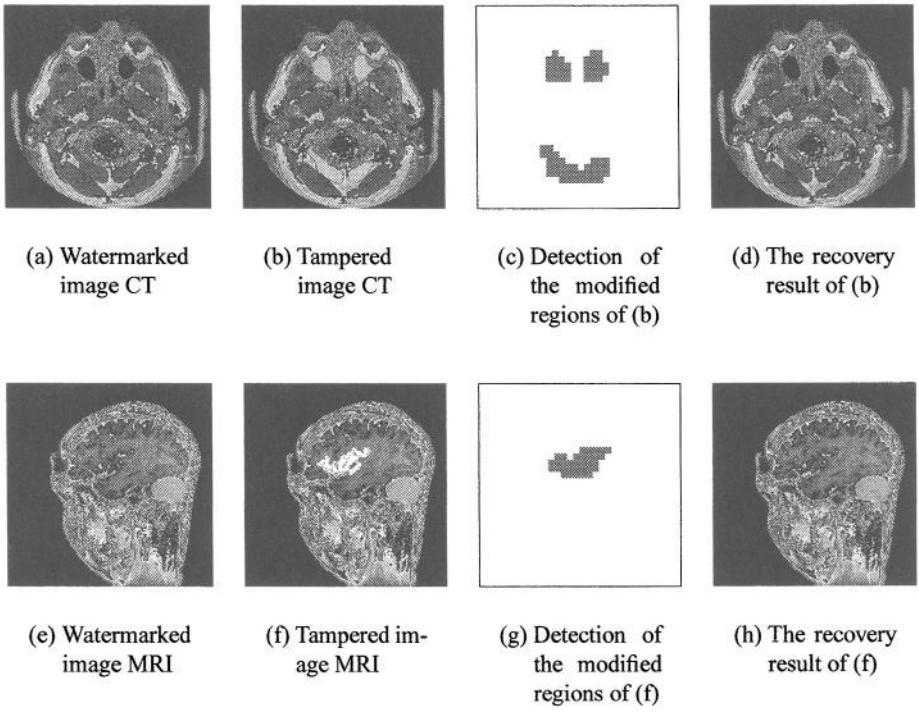


Fig. 7. CT and MRI images detection and recovery

In the following, the security analyses are shown. First, assume the attacker knows the whole methods, but does not know the DES secret key SK and a random sequence $K = \{k_0, k_1, ..., k_{\frac{N \times N}{16 \times 16}}\}$. The attacker can forge the verification data if he/she can break the DES cryptosystem. However, it is still difficult to break the DES cryptosystem from now on. Thus, the DES cryptosystem remains secure. Because the security of our scheme is based on the DES cryptosystem, it is confirmed that our scheme can

withstand the malicious attacks. On the other hand, our scheme uses different seed SD_2 to generate different sequence K for each image to prevent the counterfeit attack. Even though two image blocks are the same, without the same number k_i, their output verification data are still different from each other. In our scheme, two secret seeds SD_1 and SD_2 and one DES secret key SK are kept secretly, so it is impossible for attackers to counterfeit an image block if the correct keys are not available.

5 Conclusions

We have proposed a recoverable image tamper proofing technique using the symmetric key cyrptosystem and vector quantization for detecting and restoring of a tampered medical image. Our scheme not only can detect the modified places of the host image, but also can recover the content of the altered image. Besides, the proposed scheme can withstand the counterfeit attack. Furthermore, the watermarked image of our scheme can obtain high image quality. Therefore, our proposed scheme is very suitable for medical image authentication.

References

1. M. Holliman and N. Memon, "Counterfeiting attacks on oblivous block-wise independent invisible watermarking schemes," *IEEE Transactions on Image Processing ,* Vol. 9, No. 3, pp. 432–441, March 2000.
2. D. Kundur and D. Hatzinakos, "Digital watermarking for telltale tamper proofing and authentication," *Proceedings of the IEEE,* Vol. 87, No. 7, pp. 1167–1180, 1999.
3. C.Y. Lin and S.F. Chang, "A robust image authentication method distingulish JPEG compression from malious manipulation," *IEEE Transactions on Circuits and Systems of Video Technology,* Vol. 11, No. 2, pp. 153–168, February 2001.
4. Y. Linde, A. Buzo, and R.M. Gray, "An algorithm for vector quantizer design," *IEEE Transactions on Communications ,* vol. 28, pp. 84–95, 1980.
5. NBS, "Data Encryption Standard (DES)," *Federal Information Processing Standards Publication 46, National Technical Information Service ,* Springfield, VA, 1997.
6. R. Rivest, A. Shamir and L. Adleman, "A method for obtaining digital signatures and public-Key cryptosystems," *Communications of the ACM,* Vol. 21, No. 2, pp. 120-126, 1978.
7. P. Tsai, Y. C. Hu, and C. C. Chang, "Using set partitioning in hierarchical trees to authenticate digital images," *Signal Processing: Image Communication ,* Vol. 18, pp. 813–822, 2003.
8. P.W. Wong, "A public key watermark for image verification and authentication," *Proceedings of the IEEE International Conference on Image Processing ,* Vol. 1, pp. 455-459, Chicago, October, 1998.
9. L. Xie, G.R. Arce, and R.F. Gravman, "Approximate image message authentication codes," *IEEE Transactions on Multimedia ,* Vol. 3, No. 2, pp. 242–252, June 2001.

Efficient Lossy to Lossless Medical Image Compression Using Integer Wavelet Transform and Multiple Subband Decomposition

Li-bao Zhang[1], Ke Wang[1]

[1] College of Communication Engineering, Jilin University,
Changchun 130025, China
{ZHANG Li-bao, WANG Ke, libaozhang}@163.com

Abstract. Since medical imaging produce prohibitive amounts of data, efficient and low-complexity compression is necessary for storage and communication purposes. In this paper, a flexible coding algorithm called embedded multiple subband decomposition and set quadtree partitioning (EMSD-SQP) is presented based on integer wavelet transform (IWT). The presented method exploits three new coding strategies-multiple subbands decomposition (MSD), multiple subbands scaling (MSS) and fast quadtree partitioning (FQP). During transform, three high frequency subbands are secondly decomposed using IWT respectively for optimizing the transform coefficient distribution of high frequency subbands. Then, each subband is scaled according to its significance. The scaling factors are the integer powers of two. Finally, all image coefficients are encoded using fast quadtree partitioning scheme, which improves the lossy compression performance and reduces the memory demands. Simulation results for CT and MRI images show that the EMSD-SQP algorithm provides PSNR performance up to 4~6 dB better than SPIHT and SPECK using IWT. And the PSNR performance of EMSD-SQP has 0.4~0.8 dB better than SPIHT using Daubechies 9/7 discrete wavelet filters. Additionally, the lossless compression performance of the presented algorithm is quite competitive with other efficient compression method.

1 Introduction

Medical data (CT, MRI) are useful tools for diagnostic investigation, however their usage may be made difficult because of the amount of data to store or because of the duration of communication over a limited capacity channel. Efficient image storage and transmission is in great demand for medical community. Lossy medical image compression techniques may have the potential for widespread medical acceptance. However, the compression will reduce the image fidelity, especially when the images are compressed at lower bit rates. For example, the reconstructed images, which are coded using JPEG, can suffer from blocking artifacts and the image quality will be severely degraded under the circumstance of high compression ratios.

The discrete wavelet transform (DWT) is widely used in image compression because it avoid blocking artifacts and improve the compression performance [1-5].

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 86-93, 2004.

More recently, a new still image compression standard–JPEG2000 is proposed based on the DWT. Most of the DWT-based codecs support an embedded bitstream. This means that the quality of the reconstructed image increases when more encoded bits can be available to the decoder. However, the main drawback of the DWT is that the wavelet coefficients are floating-point numbers. In this case efficient lossless coding is not possible using DWT. But in many applications, efficient lossless medical image compression is very important for diagnostic investigation and analysis.

The lifting scheme (LS) presented by Sweldens, allows a low-complexity and efficient implementation of the DWT [2]. One such transform is the LS-based integer wavelet transform (IWT) scheme [3]. The IWT mainly has three advantages. Firstly, a lossless decoding image can be reconstructed perfectly. This point is very significant for medical image processing. Secondly, IWT has lower computational complexity than DWT. Finally, the use of IWT is a means to reduce the memory demands of the compression algorithm. However, using IWT instead of DWT degrades the performances of the lossy codecs. This is due to the fact that the transform is no more unitary, and the information content of each coefficient is no longer directly related to magnitude; this is particularly harmful for encoders with rate allocation based on bitplanes, such as SPIHT [4] and SPECK [5] coding scheme.

In this paper, we present a new coding scheme for medical image compression, called embedded multiple subband decomposition and set quadtree partitioning (EMSD-SQP). The new algorithm exploits three new coding strategies-multiple subbands decomposition (MSD), multiple subbands scaling (MSS) and fast quadtree partitioning (FQP). During transform, three high frequency subbands are secondly decomposed using IWT respectively for optimizing the transform coefficient distribution of high frequency subbands. All image coefficients are encoded using fast quadtree partitioning scheme. The new algorithm, in addition to reducing computational complexity and enhancing coding flexibility of IWT, supports efficient both lossless and lossy medical image coding using a single bitstream.

2 Integer Wavelet Transform

This LS implementation [2], [3] is a very efficient implementation of the DWT. It exploits the redundancy between the high pass (HP) and low pass (LP) filters necessary for perfect reconstruction (PR). It reduces the number of arithmetic operations up to a factor of two compared to the filter-bank (FB) implementation. Its structure guarantees that the scheme is reversible, regardless of the filters used. Its low computational complexity and efficient lossless compression performance is very useful to real-time image transmission. In [2], Sweldens and Calderbank presented the reversible integer-to–integer wavelet transforms based on the Lifting Scheme (LS). In LS, the integer wavelet transforms can be described through *polyphase matrix* using *Euclidean Algorithm* as

$$P(z) = \prod_{i=1}^{m} \begin{bmatrix} 1 & s_i(z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ t_i(z) & 1 \end{bmatrix} \begin{bmatrix} K & 0 \\ 0 & 1/K \end{bmatrix} \tag{1}$$

$P(z)$ can be defined as analysis filters. $s_i(z)$ and $t_i(z)$ can be defined as *Laurent polynomials*. In table 1, we use the notation *(x,y)* to indicate that the analyzing and synthesizing wavelet functions have x and y vanishing moments, respectively. In the forward transform equations, the input signal, lowpass subband signal, and highpass subband signal are denoted as $x[n]$, $s[n]$ and $d[n]$, respectively. For convenience, we also define the quantities $s_0[n]=x[2n]$ and $d_0[n]=x[2n+1]$.

Table 1. Several forward transform of IWTs

Name (*x*,*y*)	Forward transform of IWT
(2,2)	$\begin{cases} d[n]=d_0[n]-\lfloor 1/2(s_0[n+1]+s_0[n])\rfloor \\ s[n]=s_0[n]+\lfloor 1/4(d[n]+d[n-1])+1/2\rfloor \end{cases}$
(4,2)	$\begin{cases} d[n]=d_0[n]+\lfloor 1/16((s_0[n+2]+s_0[n-1])-9(s_0[n+1]+s_0[n])+1/2)\rfloor \\ s[n]=s_0[n]+\lfloor 1/4(d[n]+d[n-1])+1/2\rfloor \end{cases}$
(3,3)	$\begin{cases} d_1[n]=d_0[n]-s_0[n] \\ s[n]=s_0[n]+\lfloor 1/16(8d_1[n]+d_1[n-1]-d_1[n+1])+1/2\rfloor \\ d[n]=d_1[n]+\lfloor 1/16(s[n+2]-s[n-2]+6(s[n-1]-s[n+1])+1/2\rfloor \end{cases}$
(4,4)	$\begin{cases} d[n]=d_0[n]+\lfloor 1/16((s_0[n+2]+s_0[n-1])-9(s_0[n+1]+s_0[n])+1/2)\rfloor \\ s[n]=s_0[n]+\lfloor 1/32(9(d[n]+d[n-1])-(d[n+1]+d[n-2])+1/2\rfloor \end{cases}$

3 Description of EMSD-SQP Coding Algorithm

3.1 Multiple Subbands Decomposition

Integer wavelet transform has worse energy compaction than common wavelet transform, which is a disadvantage for efficient medical image compression. In order to improving the compression performance, high frequency subbands are secondly decomposed using the same IWT. The new wavelet decomposition strategy is similar to the wavelet packet decomposition. However, there are two important differences between multiple subbands decomposition and discrete wavelet packet.

- Because the integer wavelet transform is nonlinear transform, the scaling factors are necessary for all subbands to optimize the subband coefficient distribution. So general discrete wavelet packet theory is unsuitable to IWT.
- In multiple subbands, only are the HH_1, HL_1 and LH_1 subbands decomposed secondly. This is due to the fact that if more subbands are decomposed secondly, more scaling factors will be chosen and optimized. The computational and coding complexity will increase quickly, but the lossy compression performance is not improved apparently.

Fig. 1 shows the decomposed MRI image using IWT based on different methods. We adopt (2,2) IWT and the transform level number is four.

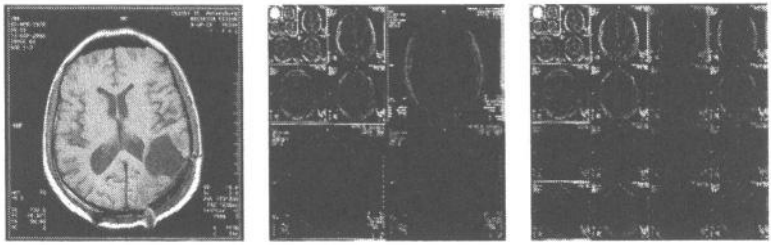


Fig. 1. Original MRI image (left) and the decomposed MRI image using IWT (middle) and the decomposed MRI image using IWT with MSD (right)

3.2 Multiple Subbands Scaling

In [7], the scaling step can be operated with multiplying the lowpass coefficients by K (scaling parameter-SP) and highpass coefficients by $1/K$ as illustrated in (1). However, for general IWT, K is the irrational number. For offering an integer-to-integer version, three extra lifting steps must be implemented instead of original scaling transform, which improves the transform complexity and increases the error of the mantissa rounding operations.

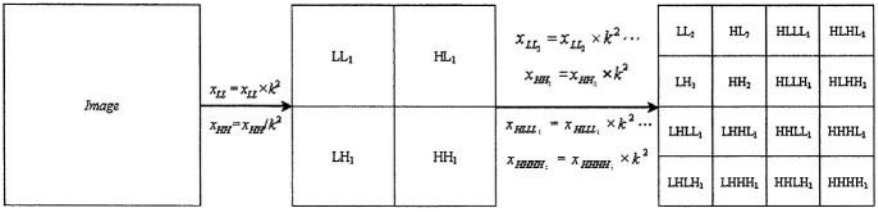


Fig. 2. Subband scaling step in an image transform

In this paper, we analyze the characterizations of IWTs and present a new method to reducing the computational complexity of IWTs. For general IWTs, the scaling parameter K is $\sqrt{2}$. If we omit the mantissa rounding operation, the coefficients of HH will be multiplied by $1/K^2$, the coefficients of LL will be multiplied by K^2, and all coefficients of HL and LH are invariant. Fig.2 shows the scaling step in a two-dimensional transform. For general IWT, the coefficients of HH will be multiplied by 1/2 and the coefficients of LL multiplied by 2. Because we use integer powers of two as quantization threshold during encoding, the scaling parameter of IWTs is exactly equal to the quantization threshold. Thus, using MSS instead of three extra lifting steps can reduces the computational complexity.

Table 2 gives the scaling factors of different IWTs for two-dimensional wavelet decomposition with MSD. Fig. 3 shows the ordering of bitplanes from $HHHH_1$ to LL_2 without scaling.

Fig. 4 presents the ordering of bitplanes from $HHHH_1$ to LL_2 with scaling presented by OSF.

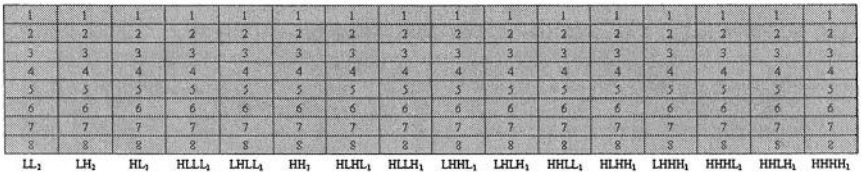


Fig. 3. Ordering of bitplanes from $HHHH_1$ to LL_2 without scaling

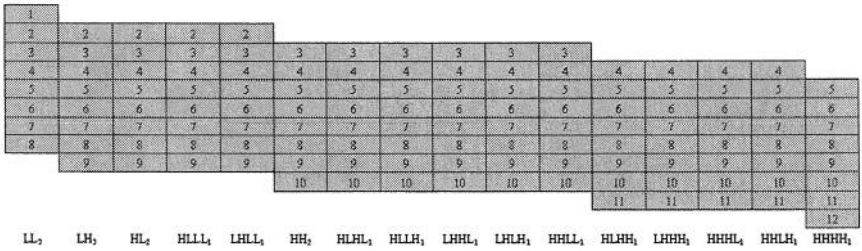


Fig. 4. Ordering of bitplanes from $HHHH_1$ to LL_2 with scaling based on MSD and OSF

Table 2. Scaling factors for different IWTs for two-dimensional wavelet decomposition

Wavelet	Subband															
	LL	HL	LH	HH	HL	HL	HL	HL	LH	LH	LH	LH	HH	HH	HH	HH
	2^2	2^2	2^2	2^2	LL	HL	LH	HH	LL	HL	LH	HH	LL	HL	LH	HH
(2,2)	4	2	2	1	2	1	1	1/2	2	1	1	1/2	1	1/2	1/2	1/4
(4,2)	4	2	2	1	2	1	1	1/2	2	1	1	1/2	1	1/2	1/2	1/4
(3,3)	4	2	2	1	2	1	1	1/2	2	1	1	1/2	1	1/2	1/2	1/4
(4,4)	4	2	2	1	2	1	1	1/2	2	1	1	1/2	1	1/2	1/2	1/4

3.3 Fast Quadtree Partitioning Scheme

All transform coefficients are coded using fast quadtree partitioning (FQP) scheme after the each subband was scaled. The coder only uses a simple quadtree partitioning, which is similar to SPECK algorithm, instead of both quadtree partitioning and the octave band partitioning of SPECK. Fig.5 shows the basic scheme of FQP algorithm.

We adopt the integer powers of two as the threshold of coefficient quantization. We say that a set Ω of coefficients is *significant* with respect to n if

$$\max_{(i,j)\in\Omega}\{|c_{i,j}|\} \geq 2^n \quad (n = 0,1,2,3\ldots) \tag{2}$$

Otherwise it is *insignificant.* We can write the significance of a set Ω as

$$\Gamma_n(\Omega) = \begin{cases} 1, & \text{if } 2^n \leq \max_{(i,j)\in\Omega}|c_{i,j}| < 2^{n+1} \\ 0 & \text{else} \end{cases} \tag{3}$$

When the outcome of the test is "1", we say the set is *significant* for bitplane *n,* otherwise it is *insignificant.*

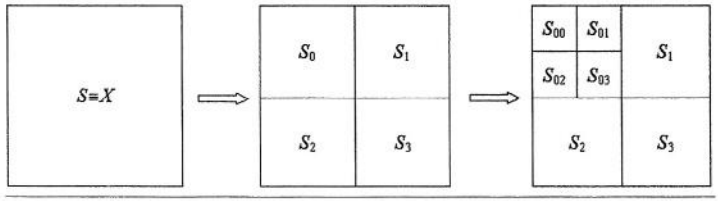


Fig. 5. Partitioning scheme of FQP algorithm

Starting with *S* as the root set from the whole map, If *S* is significant for threshold n_{max}. Set *S* is spitted into four quadrant sets, collectively denoted *O(S).* Four subsets are respectively corresponding to all subbands in the same level. We adopt the significance test for the same *n* to each of these sets and split four again only if significant. Significant sets continue to be recursively split until all there are four pixels, where upon, the significant ones are found and appended to a list of significant pixels, called the LSP. The sizes and coordinates of all insignificant sets and pixels can be appended to an array of insignificant sets and pixels (AISP). The progressing is not stop until all current sets of type *S* have been tested against *n.* Then, *n* is decremented by 1, and the FQP is continued for all Set *S* in AISP until *n* is equal to 1.

4 Experimental Results

We compare the EMSD-SQP algorithm with SPIHT and SPECK. Different medical images are selected as the test images.

Table 3 shows comparison of the PSNR performances among the EMSD-SQP algorithm, the based-IWT SPIHT and SPECK for Barbara, MRI and CT images. We adopt (3,3), which is the best IWT for compression. In table 4, the PSNR values using the EMSD-SQP algorithm based on (2,2), (4,2), (3,3) and (4,4) for MRI image are shown and compared with the results of DWT-based SPIHT and SPECK.

Table 5 shows lossless compression comparison of EMSD-SQP, SPIHT and SPECK on Barbara, MRI and CT image. Fig.6 gives the coding results of EMSD-SQP algorithm for the MRI image based on (3,3). Fig.7 shows the lossy reconstructed CT image using EMSD-SQP algorithm based on (3,3) at 0.25 bpp and 1.0bpp.

5 Conclusions

In this paper, we propose a so-called EMSD-SQP algorithm that has three primary advantages for medical image coding. Firstly, multiple subband decomposition (MSD) optimizes the transform coefficient distribution of each subband and improves energy compaction. Secondly, each subband is scaled according to its significance.

The scaling factors are the integer powers of two. Thirdly, fast quadtree partitioning (FQP) is presented to avoid octave band partitioning of SPECK, reduce the computational complexity and increases the lossy and lossless compression efficiency. We expect this idea is valuable for future research in medical image coding and its applications.

Table 3. Comparison of lossy compression results among EMSD-SQP, SPIHT and SPECK based on (3,3) IWT for *Barbara*, *MRI* and *CT* images

Image	PSNR (dB)								
	EMSD-SQP			SPIHT			SPECK		
bpp	0.25	0.5	1.0	0.25	0.5	1.0	0.25	0.5	1.0
Barbara	28.18	32.21	36.84	20.21	24.56	30.11	20.72	25.04	30.34
MRI	26.49	31.78	37.34	19.92	25.68	30.97	19.46	25.44	30.77
CT	27.19	31.48	35.93	20.67	25.81	30.63	20.11	25.27	29.97

Table 4. Comparison of lossy coding methods for *MRI* images using (2,2), (4,2), (3,3) and (4,4) and Daubechies 9/7

Filters	PSNR (dB)					
	(2,2)	(4,2)	(3,3)	(4,4)	Daubechies 9/7	
bpp	EMSD-SQP	EMSD-SQP	EMSD-SQP	EMSD-SQP	SPECK	SPIHT
0.25	25.21	25.48	26.49	25.98	25.92	25.58
0.5	30.74	31.02	31.78	31.59	31.56	31.30
1.0	36.23	36.52	37.34	37.13	37.09	36.87

Table 5. Comparison of different compression method for lossless coding (in bpp)

Methods	Image		
	Barbara	MRI image	CT image
JPEG-LS [7]	4.863	4.019	4.917
SPIHT	4.711	3.981	4.879
EMSD-SQP	4.634	3.912	4.805

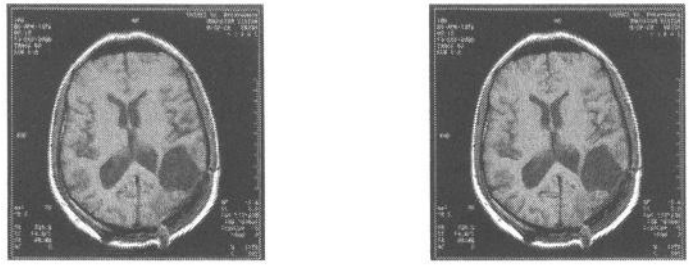


Fig. 6. Coding results of *MRI* image using EMSD-SQP algorithm based on (3,3) at 0.25 bpp (left), 1.0bpp (right).

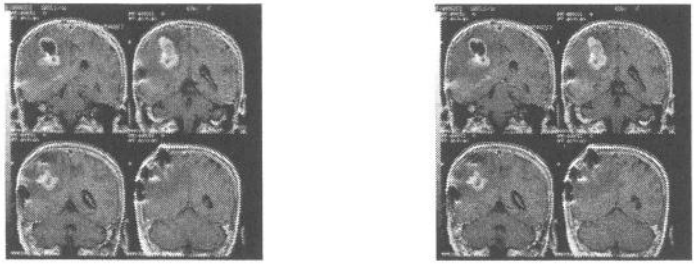


Fig. 7. Coding results of *CT* image using EMSD-SQP algorithm based on (3,3) at 0.25 bpp (left), 1.0bpp (right).

References

1. Grangetto, M., Magli, E. and Martina, M.: Optimization and implementation of the integer wavelet transform for image coding. IEEE Transaction on Image Processing. Vol. 11. No. 6. (2002) 596~604
2. Daubechies, I., Sweldens, W.: Factoring wavelet transforms into lifting steps. Journal of Fourier Analysis Application. Vol. 4. No. 3. (1998) 247~269
3. Calderbank, R.C., Daubechies, I. and Sweldens, W.: Lossless image compression using integer to integer wavelet transforms. IEEE International Conference on Image Processing. v1. (1997), 596~599
4. Said, A., Pearlman, W.A.: New Fast and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees. IEEE Transaction Circuit System Video Technology. Vol. 6. No. 3. (1996) 243-249
5. Islam, A. and Pearlman, W.A.: Embedded and efficient low-complexity hierarchical image coder. Visual Communications and Image Processing. Proceedings of SPIE. Vol. 3653. (1999), 294~305
6. Daubechies I, Sweldens W.: Factoring wavelet transforms into lifting steps. Journal of Fourier Analysis Application. Vol. 4. No. 3. (1998) 247~269.
7. Weinberger MJ, Seroussi, and Sapiro G. LOCO-1: A low complexity, context-based lossless image compression algorithm. Processing Data Compression Conference. (1996) 140~149

Geodesic Active Regions Using Non-parametric Statistical Regional Description and Their Application to Aneurysm Segmentation from CTA

Monica Hernandez and Alejandro F. Frangi

Computer Vision Group, Division of Biomedical Engineering, Aragon Institute of Engineering Research, University of Zaragoza, Zaragoza, Spain
{mhg,afrangi}@unizar.es

Abstract. The inclusion of statistical region-based information in the Geodesic Active Contours introduces robustness in the segmentation of images with weak or inhomogeneous gradient at edges. The estimation of the Probability Density Function (PDF) for each region, involves the definition of the features that characterize the image inside the different regions. PDFs are usually modelled from the intensity values using Gaussian Mixture Models. However, we argue that the use of up to second order information could provide better discrimination of the different regions than based on intensity only, as the local intensity manifold is more accurately represented. In this paper, we present a non parametric estimation technique for the PDFs of the underlying tissues present in medical images with application for the segmentation of brain aneurysms in CTA data with the Geodesic Active Regions model.

1 Introduction

Brain aneurysms are pathological dilatations of cerebral arteries developed on weakened vessel walls due to blood pressure. Two dimensional Digital substraction angiography (DSA) is considered the gold standard technique for the detection and quantification of brain aneurysms. However, other less invasive acquisition techniques like Computed Tomography Angiography (CTA), Magnetic Resonance Angiography (MRA) or 3D Rotational Angiography (3DRA) are also used as complementary methods for these aims [15, 7]. In clinical practise, quantification is usually performed from Maximum Intensity Projections (MIP) of the original volumetric scan, which introduces a high degree of subjectivity to the quantification of the aneurysm. The use of computerized 3D segmentation techniques can play a crucial role in improving quantification of the aneurysm dimensions as well as for a correct interpretation of the 3D morphology.

The adoption of deformable models for segmentation in vascular cerebral structures has become very popular over the last years [9,16,1]. In particular, the ability to handle with topological changes in complex structures makes implicit

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 94–102, 2004.

deformable models a very suitable technique for modelling the shape of vascular structures and brain aneurysms [11].

Traditionally, implicit deformable models based on the Geodesic Active Contours approach depend on the gradient of the image as edge integration criteria. Due to the low quality of the medical data, the front in evolution usually suffers from leakage in places with weak or inhomogeneous image gradient, and usually does not provide good results in brain vessels. There have been several efforts to include statistical region-based information in the process of segmentation [17,12]. In places with weak gradient, the region-based information drives the evolution of the active contour providing more robust segmentation. Previous attempts for the inclusion of statistical region-based information into the deformable model has shown promising results in the segmentation of brain aneurysms in 3DRA and CTA data [3, 8].

The inclusion of statistical region-based information into the deformable model is done using region descriptors, which are defined in terms of the negative logarithm of a Probability Density Function (PDF) associated to the region [3]. The estimation of the PDF for each region, involves the definition of the features that characterize the image inside the different regions. In fact, the estimated PDF can be considered as a conditional PDF $P(\mathbf{x}|\mathbf{f})$ where $\mathbf{x}$ is the point in the image domain and $\mathbf{f}$ is the vector of features used to describe the image in the estimation process.

In most previous attempts the estimation of the PDF for region descriptors are based on two main assumptions: image intensity is the most discriminant regional descriptor, and the statistics of image intensity can be described using parametric estimators of the PDF. In particular, PDFs are usually modelled from the intensity values using a Gaussian Mixture Model (GMM) [17, 12, 3] with parameters estimated via the Maximum Likelihood principle. However, we argue that the use of up to second order information could provide better approximations for the different regions as, besides intensity, local geometrical information is introduced for region characterization and therefore, improve the segmentation. As the use of a Gaussian model with a higher dimension feature set could generalize poorly, the use of non parametric estimation techniques becomes necessary for PDF estimation.

This article proposes a method of introducing high order information for the estimation of PDF of the different tissues present in medical data. The technique is here applied to the segmentation of brain aneurysms in CTA data. The novelty of our method stems in the use of differential image descriptors of up to second order for the definition of non-parametric region descriptors of the main tissue types that are present in the CTA images. The underlying PDFs are estimated using adaptive Parzen windows based on the k-Nearest Neighbor (kNN) rule. The result is an algorithm that improves region PDF estimation in an optimal way, as feature selection is included in the method, providing accurate segmentations.

The paper is organized as follows. Section 2 explains the devised PDF estimation method used in the segmentation with Geodesic Active Regions model. The results of the method and conclusions are reported in Section 3.

2 Probability Density Function Estimation

For the estimation of the PDF for vessel, background and bone tissues in CTA, a method based on a non parametric estimation technique is proposed. First, the K-means algorithm is used for unsupervised construction of the train set. Then, the kNN rule is used for PDF estimation in a multidimensional feature space defined by the multiscale derivatives of the image up to second order. The result is an estimation technique that takes into account not only the intensity distribution in the image but also approximates to a higher degree the local image structure.

2.1 Train Set Construction

For the construction of the train set in the learning step of the PDF estimation, we propose to use a fully automatic way to define correct labelled samples. The intensity values of the CTA image are considered as a set of observations $\{i_1, ..., i_n\}$ of a random variable I of density $P(I)$. A GMM is asumed for $P(I)$. K-means clustering is used to extract a uniform random sample of correct labelled points to build the train set from a prototype sample of images from our CTA data. This clustering correspond to the regions of the space of intensities that contains the modes of $P(I)$. The estimation of the GMM parameters is done using the Expectation Maximization (EM) algorithm. These parameters are used for initialization of the K-means algorithm. Sample images without bone are used for extracting features from vessel and background tissues. Sample images with bone are used for feature extraction from bone and background tissues. The details for practical implementation are given in subsection 2.3.

From the clustering performed by K-means method we can infer that tissue PDFs cannot be represented by a GMM as vessel and bone intensities in CTA widely overlap. Then, more sophisticated PDF estimation methods have to be designed to achieve an accurate representation of the region-based descriptors.

2.2 Probability Density Function Estimation

In our framework, the problem of PDF estimation associated to a region, is considered as the estimation of the conditional PDF $P(\mathbf{x}|\mathbf{f})$ at point $\mathbf{x}$ where $\mathbf{f}$ is the vector of features used to represent the image in the estimation process. The use of up to second order information is used to define the vector of features introducing both intensity and local geometrical information for region characterization. Due to the high dimension of the feature vector, non parametric estimation techniques are used for the estimation of the PDFs.

Image Features A common approach to analyze the local behavior of an image, $I(\mathbf{x})$, is to consider its Taylor expansion in the neighborhood of a point $\mathbf{x_0}$ at scale σ, $I_\sigma(\mathbf{x}_0 + \delta\mathbf{x}_0) \approx I_\sigma(\mathbf{x}_0) + \delta\mathbf{x}_0^T \nabla_\sigma I(\mathbf{x}_0) + \delta\mathbf{x}_0^T H_\sigma(\mathbf{x}_0)\delta\mathbf{x}_0$, where I_σ, $\nabla_\sigma I$, and H_σ are the image intensity function, the gradient vector, and the Hessian

matrix of the image computed after convolution with a Gaussian kernel of scale σ. While the intensity provides information about the tissue properties (i.e. X-ray absorption in CTA), the norm of the gradient indicates the amount of contrast between neighboring tissues and, the second order structure provided by the Hessian matrix indicates local curvature and bending [5]. We argue that all this information in a multiscale framework can be relevant in characterizing tissue properties.

Non Parametric PDF Estimation The Parzen windows method [13] is the most popular non parametric estimation technique as it is a consistent estimator of any continuous density [14]. For density functions of real variable, given sample data $x_1, ..., x_l$, Parzen's method suggests the following estimate at point x_0: $P(x_0, \gamma_l) = \frac{1}{l} \sum_{i=1}^{l} K_{\gamma_l}(x_0 - x_i)$, where K_{γ_l} is a smooth function such that $\int K_{\gamma_l}(t - x_i)dt = 1$. The k-Nearest Neighbor (kNN) estimation technique [2] can be interpreted as a Parzen approach with parameter γ_l adjusted automatically depending on the location of the point [6] (i.e. $\gamma_l = |x_0 - x_{[k]}|$ where $x_{[k]}$ is the kth nearest neighbor). The generalization to higher dimensions is straightforward. Assuming that points with similar local image structure belong to the same tissue class, a kNN density estimator can approximate the probability for a given voxel to belong to a class.

In our case, the local image structure is parameterized with a feature vector derived from the second order Taylor expansion. For a point $\hat{\mathbf{x}}$ in the train set, we associate the feature vector

$$\mathbf{f}(\hat{\mathbf{x}}) = [\mathbf{f}_{\sigma_0}, ..., \mathbf{f}_{\sigma_d}] \text{ with } \mathbf{f}_{\sigma_i}(\hat{\mathbf{x}}) = (I_{\sigma_i}, |\nabla I_{\sigma_i}|, \lambda_{1_{\sigma_i}}, \lambda_{2_{\sigma_i}}, \lambda_{3_{\sigma_i}}) \tag{1}$$

where I_{σ_i} represents the convolution of the image with a Gaussian kernel and ∇I_{σ_i} its gradient. The parameters $\lambda_{j_{\sigma_i}}$ represent the eigenvalues of the Hessian matrix of the image I_{σ_i}, ordered by increasing magnitude. The set of scales is chosen according to an exponential sampling, $\sigma_i = \sigma_0 \cdot e^{i\rho}$, as suggested by Scale Space theory.

At this point, kNN rule is used to estimate the underlying PDF as follows. For a given voxel $\mathbf{x}$, the feature vector $\mathbf{f}(\mathbf{x})$ is defined as in Equation (1). Then, the k nearest feature vectors are found in the train set according to the Euclidean distance. The probability for a voxel of intensity i to belong to a tissue class C_j, is approximated by the formula

$$P(I(\mathbf{x}) = i | C_j) = \frac{\sum_{\hat{\mathbf{x}} \in \mathcal{L}_j \cap \mathcal{N}_k(\mathbf{x})} K_\gamma(\mathbf{f}(\mathbf{x}), \mathbf{f}(\hat{\mathbf{x}}))}{\sum_{\hat{\mathbf{x}} \in \mathcal{N}_k(\mathbf{x})} K_\gamma(\mathbf{f}(\mathbf{x}), \mathbf{f}(\hat{\mathbf{x}}))} \tag{2}$$

where $\mathcal{L}_j$ represents the set of points in the train set that belongs to the class C_j, $\mathcal{N}_k(\mathbf{x})$ is the set of the k nearest neighbors and K_γ is the Gaussian kernel with standard deviation equal to the Euclidean distance to the kth nearest neighbor.

2.3 Details of Implementation

Train Set Construction and Image Features Six sample images were selected from the whole CTA data base to build the train set. Three of the images did not present bone tissue in the Kmeans classification. The vessels present in these images roughly covered the range of widths in the data base. The aneurysms covered also representative dome sizes. The three images that presented bone tissue included also typical bony structures. A total of 9000 points were randomly selected for each tissue, resulting on a total of 27000 points.

The minimum scale was selected equal to the in-plane voxel size (i.e. $\sigma_0 = 0.4$ mm), with the number of scales, d, equal to 10, and ρ equal to 0.2. This way, the maximum scale considered was 2.95 mm which covers approximately objects from 0.8 to 6 mm, this is, from the thinnest arteries to the mean dome size present in our aneurysm data base. As it is customary in pattern recognition theory, the feature vectors of the train and test sets were normalized [4].

Optimal Number of Neighbors and Feature Selection Cross-Validation (CV) was used for estimating the prediction error of the kNN rule as well as for the selection of the optimal number of neighbors, k [6]. As loss function, we used the cross-entropy loss $L(C_j, \widehat{C_j}(\mathbf{x})) = -2\sum_{k=0}^{2}\delta(C_j = \widehat{C_j}(\mathbf{x}))\log(P(I(\mathbf{x}) = i|C_j))$, where C_j corresponds to a tissue class, $\widehat{C_j}(\mathbf{x})$ corresponds to the label of $\mathbf{x}$ in the train set, and δ corresponds to Dirac's delta function.

The prediction error, is approximated by the K-fold cross validation estimate $CV(k) = \frac{1}{N}\sum_{i=1}^{N} L(C_j, \widehat{C_j}(\mathbf{x_i}), k)$ where N is the number of points in the train set and k is the number of neighbors used to estimate $P(I(\mathbf{x}) = i|C_j)$. A prediction error study was performed over the train set using five folds. Figure 1(a) plots the mean prediction error curve performed in 10 different experiments in a range of neighbors from 1 to 500. All the cross validation curves are similar to the mean curve. The mean minimum value is reached at neighbor 30. So we considered the use of 30 neighbors as optimal choice for the generation of the PDF.

To improve the performance of the classifier and deal with the curse of dimensionality, an study for the selection of the optimal features was carried out following the guidelines given in [10]. As criterion function for evaluating the goodness of the classifier we chose $J(X) = \frac{1}{1+CV}$, where X is the set of selected features in each stage of the feature selection algorithm and CV corresponds to the error estimated by 5-fold cross validation.

For the study of the monotonicity of the criterion curve, we performed a preliminary feature selection using a sequential search algorithm. Both Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS) were considered. As can be appreciated from Figure 1(b), the criterion curve has in both cases an approximated monotonic behavior if we ignore the degradation in performance when almost all features are excluded. Taking into account this non monotonicity and the number of features, we decided to use a Sequential Floating algorithm for feature selection. Both Sequential Forward Floating Selection

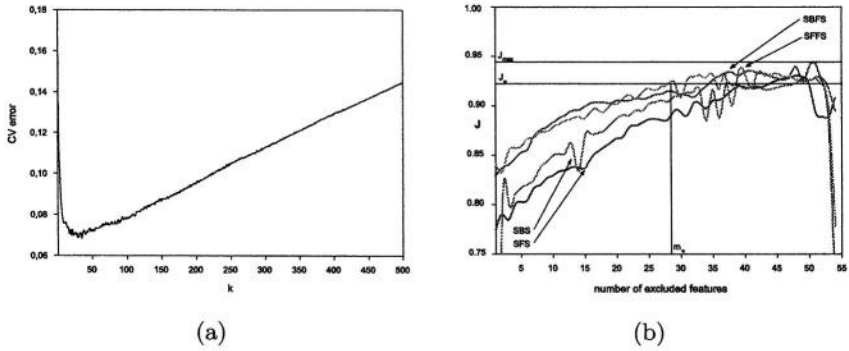


Fig. 1. (a) Mean five-fold cross validation error from the study over 10 experiments. (b) Criterion curves from sequential feature selection algorithms. The values of J_{max}, J_α and m_α are indicated on the plot.

(SFFS) and Sequential Backward Floating Selection (SBFS) were considered. As happened with the non floating sequential method, backward algorithms had better performance. For this reason, we used SBFS for feature selection. We chose the degree of degradation compared with the maximum criterion value α equal to 1%. From this α-degradation, we determined the criterion value J_α as a threshold and the corresponding number of features m_α as desired feature set. Therefore, the dimensionality of the feature set was reduced from 55 to 28. The index of overlapping for selected features performed between SBFS and SFFS was approximately equal to 81%.

3 Results and Conclusions

3.1 Results

We computed the estimated PDF with a GMM and a non parametric model and performed both segmentations in a data base of 39 brain aneurysms. Figure 2 presents two examples of the estimated PDF using the GMM and the non parametric assumption. Figure 3 presents two examples of brain aneurysm segmented with the Geodesic Active Regions method with PDF estimated using the GMM and the non parametric assumption.

3.2 Conclusions

In this work, we have presented a method for the estimation of the underlying PDF of the tissues presented on cerebral Computed Tomography Angiography images. The K-means clustering method is used for an automatic construction of the train set and a Parzen Windows estimator based on the kNN rule is used for computing the PDFs. The features used for estimation represent not only

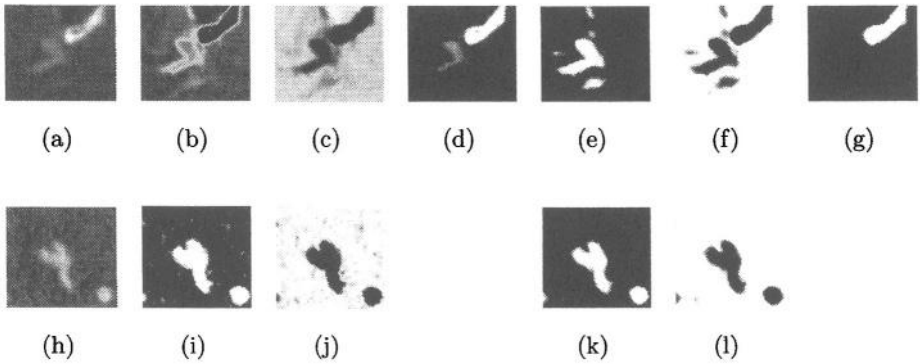


Fig. 2. PDF estimation form vessel, background and bone tissues. Light pixel values indicates high probability. The first row shows an example of brain aneurysm next to bone tissue and the second row an example without bone tissue. In this case, there is no estimated PDF for bone tissue. (a) and (h) show an slice of the grey level image. (b),(c),(d) and (i),(h) are the estimated PDF using a GMM. (e), (f), (g) and (k),(l) represent the results of the non parametric estimation.

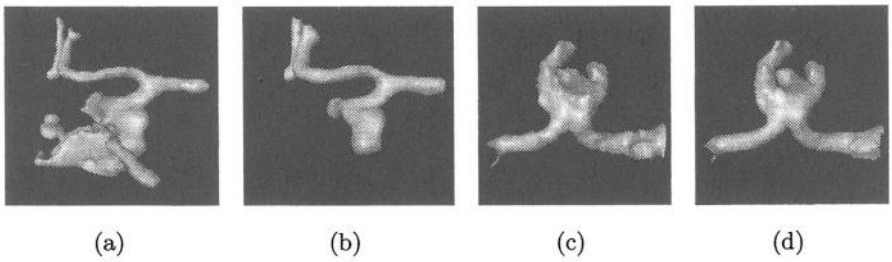


Fig. 3. Segmentation of brain aneurysms with the Geodesic Active Regions model. (a) and (b) show a model of a Posterior Communicating Artery aneurysm segmented using the GMM and the non parametric estimation methods. Bone tissue is next to the aneurysm in this example. (c) and (d) show a model of a Middle Cerebral Artery aneurysm. In this case, bone tissue is not present in the image.

the information of the intensities of the tissues but also geometrical properties of the tissues in a multiscale fashion. We believe that, due to the generality of the framework, this method could be also applied to the segmentation of other organs and/or medical imaging modalities.

The estimated PDF from the GMM show a worse performance in the examples than the ones estimated from the non parametric method, particularly, in images with bone tissue next to the aneurysms. In fact, the GMM shows a high

value of the probability for vessel tissue in the boundaries of the bone tissue thus being unable to discriminate the overlapped intensities. This model also shows high value for probability of bone in the dome of the aneurysm. The use of a feature space that discriminates between local geometrical structure improves the discrimination in this problematic locations. Starting from the same initialization, the Geodesic Active Regions achieved results that seem to be more accurate than the ones obtained with the non parametric descriptors. In images with bone tissue next to the aneurysm, the GMM based descriptors forced the front evolving towards missclassified areas recovering the part of the bone tissue coupled to the aneurysm.

In summary, we have presented a method for the estimation of tissue PDF in medical images and demonstrate that it improves the performance of the Geodesic Active Regions method in the segmentation of brain aneurysms in CTA images.

Acknowledgments: The authors would like to acknowledge to Dr. R. Barrena from the Aragon Institute of Health Sciences (Zaragoza, Spain) for providing the data and clinical background. MHG is supported by the grant AP2001-1678 from the Spanish Ministry of Education, Culture and Sport. AFF is supported by a Ramon y Cajal Research Fellowship from the Spanish Ministry of Science and Technology (MCyT). This research was partially supported by the grants TIC2000-1635-C04-01 and TIC2002-04495-C02 from the MCyT.

References

[1] S.R. Aylward and E. Bullitt, *Initialization, noise, singularities, and scale in height ridge traversal for tubular object centerline extraction,* Medical Imaging, IEEE Transactions on **21(2)** (2002 Feb), 61 – 75.

[2] B. Dasarathy, *Nearest neighbor pattern classification techniques,* IEEE Computer Society Press, Los Alamitos, CA (1990).

[3] T. Deschamps, *Curve and Shape Extraction with Minimal Path and Level-Sets techniques. Applications to 3D Medical Imaging,* Ph.D. Thesis, University of Paris-Dauphine, France (2001 Dec).

[4] R.O. Duda, P.E. Hart, and D.G.Stork, *Pattern classification,* J. Wiley (2001).

[5] A.F. Frangi, W.J. Niessen, K.L. Vincken, and M.A. Viergever, *Multiscale vessel enhancement filtering,* Lecture Notes in Computer Science **1496** (1998), 130–137.

[6] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning. data mining, inference and prediction,* Springer-Verlag (2001).

[7] A. Hochmuth, U. Spetzger, and M. Schumacher, *Comparison of three-dimensional rotational angiography with digital subtraction angiography in the assessment of ruptured cerebral aneurysms,* AJNR Am J Neuroradiol. **23(7)** (Aug 2002), 1199 – 1205.

[8] M. Holtzman-Gazit, D. Goldsher, and R. Kimmel, *Hierarchical segmentation of thin structures in volumetric medical images,* Medical Image Computing and Computer-Assisted Intervention (MICCAI 2003), Lecture Notes in Computer Science **2** (2003), 562 – 569.

[9] K. Krissian, G. Malandain, N. Ayache, R. Vaillant, and Y. Trousset, *Model based detection of tubular structures in 3d images,* Computer Vision and Image Understanding **80 (2)** (2000 Nov), 130 – 171.

[10] M. Kudo and J. Sklansky, *Comparison of algorithms that select features for pattern classifiers,* Pattern Recognition **33** (2000), 25 – 41.

[11] L.M. Lorigo, O. Faugeras, W.E.L. Grimson, R. Keriven, R. Kikinis, A. Nabavi, and C.F. Westin, *Codimension-two geodesic active contours for the segmentation of tubular structures,* Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on **1** (2000 June), 444 – 451.

[12] N. Paragios, *Geodesic active regions and level set methods: Contributions and applications in artificial vision,* Ph.D. Thesis, University of Nice Sophia-Antipolis, France (2000).

[13] E. Parzen, *On estimation of a probability density function and mode,* Ann. Math. Stats. **33** (1962), 1065 – 1976.

[14] V.N. Vapnik, *Statistical learning theory,* J. Wiley (1998).

[15] P.M. White, E.M. Teasdale, J.M. Wardlaw, and V. Easton, *Intracranial aneurysms: CT angiography and MR angiography for detection prospective blinded comparison in a large patient cohort.,* Radiology **219(3)** (2001 Jun), 739–49.

[16] P.J. Yim, J.R. Cebral, R. Mullick, H.B. Marcos, and P.L. Choyke, *Vessel surface reconstruction with a tubular deformable model,* Medical Imaging, IEEE Transactions on **20(12)** (2001), 1411 – 1421.

[17] S.C. Zhu and A. Yuille, *Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation,* Pattern Analysis and Machine Intelligence, IEEE Transactions on **18 (9)** (1996 Sept), 884 – 900.

An Efficient Method for Deformable Segmentation of 3D US Prostate Images

Yiqiang Zhan[1,2,3], Dinggang Shen[1,2]

[1] Sect. of Biomedical Image Analysis, Dept. of Radiology,
University of Pennsylvania, Philadelphia, PA
dinggang.shen@uphs.upenn.edu
[2] Center for Computer-Integrated Surgical Systems and Technology,
Johns Hopkins University, Baltimore, MD
[3] Dept. of Computer Science, Johns Hopkins University, Baltimore, MD
yzhan@cs.jhu.edu

Abstract. We previously proposed a deformable model for automatic and accurate segmentation of prostate boundary from 3D ultrasound (US) images by matching both prostate shapes and tissue textures in US images[6]. Textures were characterized by a Gabor filter bank and further classified by support vector machines (SVM), in order to discriminate the prostate boundary from the US images. However, the step of tissue texture characterization and classification is very slow, which impedes the future applications of the proposed approach in clinic applications. To overcome this limitation, we firstly implement it in a 3-level multi-resolution framework, and then replace the step of SVM-based tissue classification and boundary identification by a Zernike moment-based edge detector in both low and middle resolutions, for fast capturing boundary information. In the high resolution, the step of SVM-based tissue classification and boundary identification is still kept for more accurate segmentation. However, SVM is extremely slow for tissue classification as it usually needs a large number of support vectors to construct a complicated separation hypersurface, due to the high overlay of texture features of prostate and non-prostate tissues in US images. To increase the efficiency of SVM, a new SVM training method is designed by effectively reducing the number of support vectors. Experimental results show that the proposed method is 10 times faster than the previous one, yet without losing any segmentation accuracy.

1 Introduction

Prostate cancer continues to be the second-leading cause of cancer death in American men [1]. As transrectal ultrasound (TRUS) images have been widely used for the diagnosis and treatment of prostate cancer, the accurate segmentation of the prostate from TRUS images plays an important role in many clinical applications [1]. Accordingly, a number of automatic or semi-automatic segmentation methods have been proposed. Ghanei et.al. [2] and Hu et.al. [3] designed 3D discrete deformable models to semi-automatically outline the prostate boundaries. Shao et al [4] proposed a level set method to detect the prostate in the 3D TRUS images. Gong et al [5] provided a

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 103-112, 2004.

Bayesian segmentation algorithm, based on deformable superellipses model, to segment 2D prostate contours.

We previously proposed a statistical shape model to segment the prostate from 3D TRUS images by matching both prostate shapes and tissue textures in TRUS images [6]. The effectiveness of our previous method is mainly resulted from the joint use of two novel techniques, (i) a Gabor filter bank used for 3D texture features extraction and (ii) support vector machines (SVM) used for texture-based tissue classification. However, both of these two techniques are computationally very expensive, thereby impeding the fast segmentation of prostates from 3D TRUS images. To overcome this limitation, this paper presents an efficient segmentation approach, which is implemented in a 3-level multi-resolution framework and is further speeded up by two techniques respectively designed for different resolutions. In both low and middle resolutions, a Zernike moment-based edge detector is used to replace the step of SVM-based tissue classification and boundary identification, for fast boundary detection. In the high resolution, a new SVM training method is designed to improve the efficiency of SVMs by reducing the number of support vectors, which are initially required to construct a very complicated separation hypersurface for the classification of the highly confounding prostate and non-prostate tissues in TRUS images. By using these techniques, our approach for prostate segmentation is highly speeded up, yet without losing any segmentation accuracy.

2 Methods

Our previous deformable shape model [6] uses both statistical shape information and image texture information to segment the prostate boundary. Its success in prostate segmentation results from the texture analysis, which distinguishes prostate and non-prostate tissues from noisy TRUS images. However, the two techniques employed in texture analysis, i.e., a Gabor filter bank for texture characterization and SVMs for texture-based tissue classification, are both computationally very expensive. For example, it takes about 40 minutes to segment a prostate from a 256x256x176 TRUS image, using SGI workstation with a 500MHz processor. Therefore, it's necessary to speed up the segmentation approach.

For fast segmentation, we firstly formulate the segmentation approach in a 3-level multi-resolution framework, which has been widely used to increase the speed of the algorithms in the literature [7,8]. For example, the original TRUS image is decomposed into three multi-resolution images, i.e. the image of original size and the images down-sampled by factors 2 and 4. The surface model is initialized at the lowest resolution, and subsequently deforms to the prostate boundary. The segmentation result in the lower resolution is up-sampled to the next higher resolution, and used as initialization of the deformable model in the higher resolution. These steps are iterated until the deformable model converges to the prostate boundary in the highest resolution.

Besides the multi-resolution framework designed above, two effective methods, i.e., Zernike moment-based edge detector and a new training method for generating efficient SVMs, are particularly designed to speed up the segmentation approach in

low resolutions and high resolution, respectively. The details of these two techniques are described next.

2.1 Zernike Moment Based Edge Detection

As discussed above, although a Gabor filter bank [9] is capable of extracting robust and rich texture features, it is computationally very expensive due to the use of Gabor filters at multiple scales and orientations. Additionally, as texture features are region-based features, prostate tissues in the down-sampled TRUS images usually have less distinguished texture features, compared to those in the original images. Therefore, boundary information, directly computed from the intensities, is better than texture information for guiding the deformable model in both low and middle resolutions.

Zernike moment-based edge detector has been proposed in [10]. It has three advantages in edge detection. First, as Zernike moments are integral-based operators, it is noise tolerant, which is especially important for detecting prostate boundaries in the noisy TRUS images. Second, as detailed next, this edge detection method provides a more complete description of the detected edges than the traditional edge detector, e.g. Canny edge detector. Third, as only three masks, i.e., two real masks and one complex mask, are required to get the edge features of each voxel, it is computationally more efficient than the Gabor filter bank which used 10 masks [6].

Zernike moment operator projects the image data onto a set of complex polynomials, which form a complete orthogonal set over the interior of a unit circle. For an image $f(x, y)$, its Zernike moment of order n and repetition m can be defined as:

$$Z_{nm} = \frac{n+1}{\pi} \iint_{x^2+y^2 \le 1} f(x,y) V_{nm}^{*}(\rho,\theta)\, dx\, dy \tag{1}$$

where $V_{nm}(\rho,\theta) = R_{nm}(\rho) e^{jm\theta}$

$R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} [(-1)^s (n-s)! \rho^{n-2s}] \Big/ [s!(\frac{n+|m|}{2}-s)!(\frac{n-|m|}{2}-s)!]$, and (ρ,θ) are the polar coordinates of (x,y).

Considering an ideal step edge (c.f. Fig 1), its important features include the step height k, the background gray level h, the perpendicular distance from the center of the circular kernel l, and the angle of edge with respect to the x-axis φ. All these features can be mathematically represented by three low order Zernike moments (Z_{00}, Z_{11}, Z_{20}) as:

$$\varphi = \tan^{-1}[\mathrm{Im}(Z_{11})/\mathrm{Re}(Z_{11})] \tag{2a}$$

$$l = Z_{20}/Z'_{11} \tag{2b}$$

$$k = 3Z'_{11}/2(1-l^2)^{3/2} \tag{2c}$$

$$h = \left(Z_{00} - k\pi/2 + k\sin^{-1}(l) + kl\sqrt{1-l^2}\right)/\pi \tag{2d}$$

where $Z'_{11} = Z_{11}e^{j\varphi}$, and Re(.) and Im(.) represent the real part and the imaginary part of a complex value, respectively. Similarly, Zernike moments can be used to measure the general edges in the 2D image by using eq. (2).

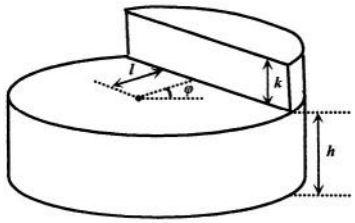


Fig. 1. A 2D ideal step edge model.

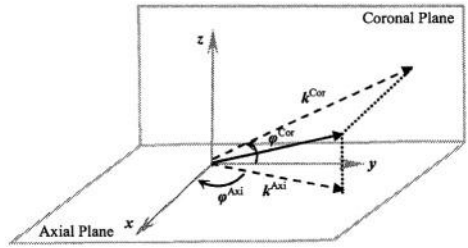


Fig. 2. Schematic explanation of using two 2D edge vectors to roughly reconstruct a 3D edge vector.

Notably, rather than extending the 2D Zernike moment to 3D, we simply apply two orthogonal 2D Zernike moment operators, which respectively lie on the axial plane and the coronal plane (c.f. Fig 2), to get two sets of edge features for each voxel *v*, i.e., $\{h^{\mathrm{Axi}}(\vec{v}), k^{\mathrm{Axi}}(\vec{v}), l^{\mathrm{Axi}}(\vec{v}), \varphi^{\mathrm{Axi}}(\vec{v})\}$ and $\{h^{\mathrm{Cor}}(\vec{v}), k^{\mathrm{Cor}}(\vec{v}), l^{\mathrm{Cor}}(\vec{v}), \varphi^{\mathrm{Cor}}(\vec{v})\}$. As shown in Fig 2, two 2D edge vectors can be roughly considered as two projections (black dashed arrows) of a 3D edge vector (black solid arrow) in the axial and the coronal planes, respectively. (Edge vector is a vector whose magnitude and direction represent the edge strength and the normal direction of the edge, respectively.) Thus, the 3D edge vector, i.e., $e(v)$, can be represented by two 2D edge vectors as follows:

$$\vec{e}(\vec{v}) \simeq k^{\mathrm{Axi}}(\vec{v}) \cdot [\cos(\varphi^{\mathrm{Axi}}(\vec{v})), \sin(\varphi^{\mathrm{Axi}}(\vec{v})), \sin(\varphi^{\mathrm{Axi}}(\vec{v}))\tan(\varphi^{\mathrm{Cor}}(\vec{v}))]^T \quad \textbf{(3)}$$

In our previous segmentation approach [6], an energy function is defined on each vertex P_i of the deformable surface model, and it is used to evaluate the matching degree of the deformable model with the prostate boundaries in the TRUS images. The energy function consists of two energy terms, i.e., the external energy, which drives the deformable model to the prostate boundary, and the internal energy, which preserves the geometric regulation of the model during deformation. By jointly minimizing these two terms, the deformable model is able to converge to the prostate boundaries. In this study, the external energy is re-formulated such that the edge features captured by Zernike moments are employed to guide the deformable segmentation, while the internal energy remains the same. Accordingly, for each vertex P_i, its external energy is defined as:

$$E^{\mathrm{Ext}}(\vec{P}_i) = w_{\mathrm{Str}} E_{\mathrm{Str}}(\vec{P}_i) + w_{\mathrm{Dist}} E_{\mathrm{Dist}}(\vec{P}_i) + w_{\mathrm{Int}} E_{\mathrm{Int}}(\vec{P}_i) \quad \textbf{(4)}$$

where $E_{\mathrm{Str}}(\vec{P}_i) = -\left(\sum_{\forall \vec{v} \in N(\vec{P}_i)} < \vec{n}(\vec{P}_i), \vec{e}(\vec{v}) > \Big/ \sum_{\forall \vec{v} \in N(\vec{P}_i)} 1 \right)$, $E_{\mathrm{Dist}}(\vec{P}_i) = \min(l^{\mathrm{Axi}}(\vec{P}_i), l^{\mathrm{Cor}}(\vec{P}_i))$ and

$$E_{\mathrm{Int}}(\vec{P}_i) = \left| \sum_{\forall \vec{v} \in N(\vec{P}_i)} (h^{\mathrm{Axi}}(\vec{v}) + h^{\mathrm{Cor}}(\vec{v})) \Big/ \sum_{\forall \vec{v} \in N(\vec{P}_i)} 2 - H(\vec{P}_i) \right|$$

There are three items in Eq. (4). The first item denotes the integrated edge strength in the spherical neighborhood of P_i, i.e., $\mathbf{N}(P_i)$. Notably, the edge strength is projected to the normal direction of the deformable surface at P_i, $n(P_i)$, by the inner product $< \cdot >$. The second item denotes the distance from P_i to the boundary. The third item requires that the deformable model converges only to the boundary with the intensity similar to the learned average intensity, $H(P_i)$, which is captured for vertex P_i from a set of training samples. By jointly minimizing these three items, the deformable model is thus driven to the prostate boundaries in both low and middle resolutions of TRUS images.

2.2 A Training Method for Increasing the Efficiency of SVM

Zernike moment-based edge detector is able to detect prostate boundaries in the low and the middle resolutions, however, it is not effective in accurately delineating prostate boundaries in the high resolution as the prostate boundaries are usually blurred by speckle noise in the original TRUS images. Accordingly, we still use the statistical texture matching method [6], which consists of texture characterization by a Gabor filter bank and texture-based tissue classification by SVMs, for prostate segmentation in the high resolution stage of our multi-resolution framework.

In our method, a set of SVMs are employed for texture-based tissue classification [6]. Each of them is attached to a sub-surface of the model surface and trained by the manually-labeled prostate and non-prostate samples around that sub-surface. In the testing stage, the input of the SVM is a feature vector, which consists of Gabor features extracted from the neighborhood of a voxel, and the output denotes the likelihood of the voxel belonging to the prostate. In this way, the prostate tissues are differentiated from the surrounding ones. However, since the Gabor features of TRUS prostate images vary greatly across the individuals and their distribution is highly overlapped between prostate and non-prostate regions, the trained SVM usually has a huge number of support vectors. This is because (i) a large number of the support vectors, locating at the margins, are required to construct a highly convoluted hypersurface, in order to separate two classes; (ii) even the highly convoluted separation hypersurface has been con-structed, quite a lot of confounding samples are still misclassified and thus selected as other support vectors, locating beyond the margins. Notably, this huge number of support vectors will dramatically increase the computational cost of the SVM. Therefore, it is necessary to design a training method to decrease the number of support vectors of the finally trained SVM, by simplifying the shape of the separation hypersurface.

The basic idea of this training method is to selectively exclude some training samples, thereby the remaining samples are possible to be separated by a less convoluted hypersurface. Since the support vectors determine the shape of the separation hypersurface, they are the best candidates to be excluded from the training set, in order to simplify the shape of the separation hypersurface.

However, excluding different sets of support vectors from the training set will lead to different simplifications of the separation hypersurface. Fig 3 presents a schematic example in the 2-dimensional feature space, where we assume support vectors exactly locating on the margins. As shown in Fig 3(a), SVM trained by all the samples has 10

support vectors, and the separation hypersurface is convoluted. Respective exclusion of two support vectors, SV_1 and SV_2, denoted as gray crosses in Fig 3(a), will lead to different separation hypersurfaces as shown in Figs 3(b) and 3(c), respectively. SVM in Fig 3(b) has only 7 support vectors, and its hypersurface is less convoluted, after re-training SVM with all samples except SV_1. Importantly, two additional samples, denoted as dashed circle/cross, were previously selected as support vectors in Fig 3(a), but they are no longer selected as support vectors in Fig 3(b). In contrast, SVM in Fig 3(c) still has 9 support vectors, and the hypersurface is very similar to that in Fig 3(a), even SV_2 has been excluded from the training set.

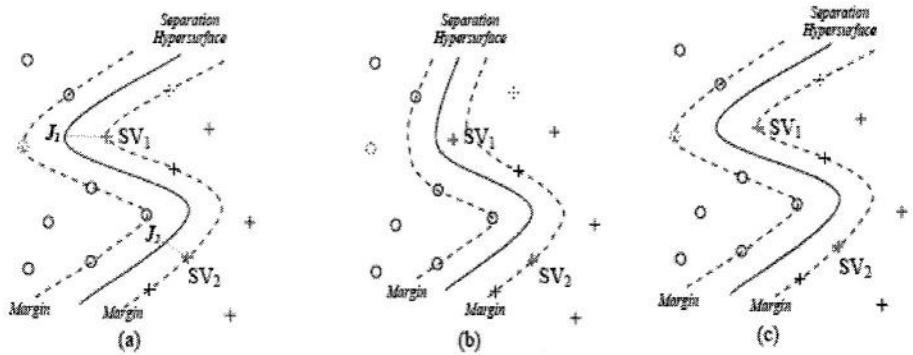


Fig.3. Schematic explanation of how to selectively exclude the support vectors from the training set, in order to effectively simplify the separation hypersurface. The solid and dashed curves denote the separation hypersurfaces and their margins, respectively. The circles and the crosses denote the positive and the negative training samples, which are identical in (a), (b) and (c). The training samples locating on the margins are the support vectors.

The reason of SVM in Fig 3(b) being more efficient than that in Fig 3(c) is that the excluded support vectors SV_1 contributes more to the convolution of the hypersurface. For each support vector, its contribution to the convolution of hypersurface can be approximately defined as the generalized curvature of its projection point on the hypersurface. For example, for SV_1 and SV_2 in Fig 3(a), their projection points on the hypersurface are J_1 and J_2. The curvature of the hypersurface at point J_1 is much larger than that at point J_2, which means the support vector SV_1 has more contribution to make the hypersurface convoluted. Therefore, it is more effective to "flatten" the separation hypersurface by excluding the support vectors, like SV_1, with their projection points having the larger curvatures on the hypersurface.

Accordingly, the new training method is designed to have the following four steps.

Step 1. Use all the training samples to train an initial SVM, resulting in l_1 initial support vectors $\{SV_i^{\text{In}}, i = 1,2,...,l_1\}$ and the corresponding decision function d_1.

Step 2. Exclude the support vectors, whose projections on the hypersurface have the largest curvatures, from the training set:

2a. For each support vector SV_i^{In}, find its projection on the hypersurface, $P(SV_i^{\text{In}})$, along the gradient of distance function d_1.

2b. For each support vector SV_i^{In}, calculate the generalized curvature of $P(SV_i^{\text{In}})$ on the hypersurface, $c(SV_i^{\text{In}})$.

2c. Sort SV_i^{In} in the decrease order of $c(SV_i^{\text{In}})$, and exclude the top n percentage of support vectors from the training set.

Step 3. Use the remaining samples to retrain the SVM, resulting in l_2 support vectors $\{SV_i^{\text{Re}}, i = 1,2,...,l_2\}$ and the corresponding decision function d_2. Notably, l_2 is usually less than l_1.

Step 4. Use the l_2 pairs of data points $\{SV_i^{\text{Re}}, d_2(SV_i^{\text{Re}})\}$ to finally train the SVRM (Support Vector Regression Machine) [12], resulting in l_3 final support vectors $\{SV_i^{\text{Fl}}, i = 1,2,\cdots l_3\}$ and the corresponding decision function d_3. Notably, l_3 is usually less than l_2.

Using this four-step training algorithm, the efficiency of the trained SVMs will be highly enhanced with very limited loss of classification rate, which will be shown in the first experiment. Notably, as in the statistical texture matching method, the matching degree of the deformable model with the prostate boundaries is defined in a noise tolerant fashion [6], a little loss of classification, i.e., a little number of mis-classfied voxels, will not influence the segmentation accuracy, while the segmentation speed is greatly increased.

3 Experimental Results

The first experiment is presented to test the performance of the proposed training method in increasing the efficiency of SVMs. We firstly select prostate and non-prostate samples from six manually labeled TRUS images. 3621 samples from one image are used as testing samples, while 18105 samples from other five images are used as training samples. Each sample has 10 texture features, extracted by a Gabor filter bank [9]. We use our method to train a series of SVMs by excluding different percentages of support vectors in Step 2c of our training method. The performances of these SVMs are measured by the number of support vectors finally used and the number of correct classifications among 3621 testing samples. As shown in Fig 4(a), after excluding 50% of initially selected support vectors, the finally-trained SVM has 1330 support vectors, which is only 48% of the support vectors (2748) initially selected in the original SVM; but its classification rate still reaches 95.39%. Compared to 96.02% classification rate achieved by original SVM with 2748 support vectors, the loss of classification rate is relatively trivial. If we want to further reduce the computational cost, we can exclude 90% of initially selected support vectors from the training set. Our finally-trained SVM has only 825 support vectors, which means the speed is triple, and it still has 93.62% classification rate. To further validate the effect of our trained SVM in prostate segmentation, the SVM with 825 support vectors (denoted by the white triangle in Fig 4(a)) is applied to a real TRUS image for tissue classification. As shown in Figs 4(b1) and 4(b2), the result of our trained SVM is not inferior to that of the original SVM with 2748 support vectors (denoted by the white square in Fig 4(a)), in terms of differentiating prostate tissues from the surrounding ones.

In the second experiment, the proposed segmentation approach is applied to segment prostates from six real 3D TRUS images. A leave-one-out validation method is used, i.e., each time five images are used for training, and the remaining one is used for testing. The size of 3D images is 256x256x176, with the spatial resolution 0.306mm. Fig 5(a) shows the multi-resolution deformation procedure on one of the TRUS im-ages. The white contours, labeled as "LF", "MF" and "HF", denote the finally de-formed models in the low, middle and high images, respectively. Notably, the models in both low and middle resolutions are guided by the Zernike moment-based edge detector, while the model in the high resolution is guided by the statistical texture matching method. The algorithm-based segmentation result is compared to the hand-labeled result in Fig 5(b). Moreover, Table 1 gives a quantitative evaluation of this comparison to all the six TRUS images. From both visual results and quantitative analysis, we can conclude that our automated segmentation method is able to segment the prostate from noisy TRUS images. Importantly, using a SGI workstation with 500MHz processor, the average running time for segmenting a prostate is 4 minutes, which is 10 times faster than our previous method [6].

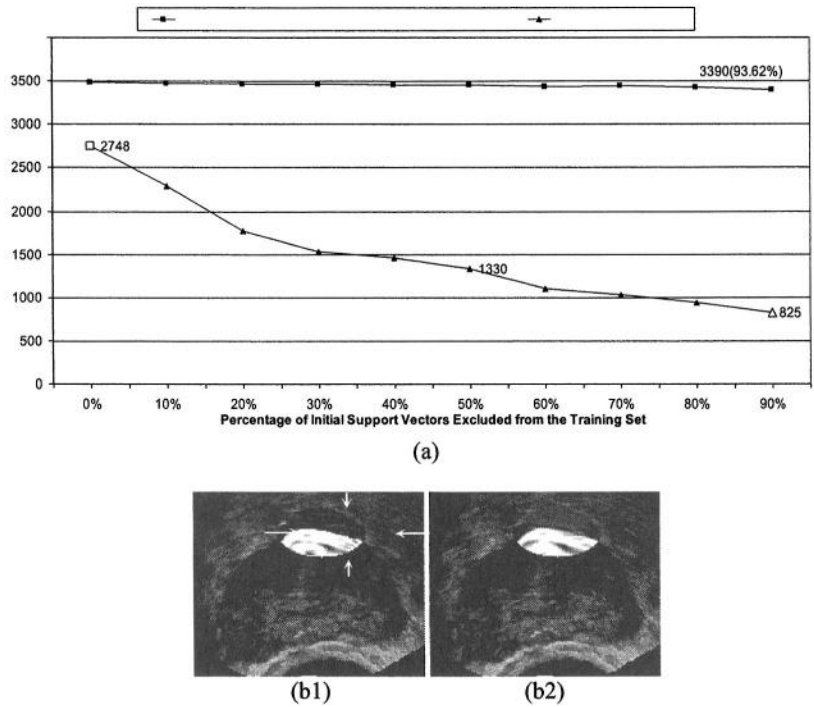


Fig. 4. (a) The performance of the finally-trained SVM changes with the percentages of initial support vectors excluded from the training set. (b) Comparisons of tissue classification results using (b1) the original SVM with 2748 support vectors and (b2) our trained SVM with 825 support vectors. The tissue classification results are shown only in an ellipsoidal region and mapped to 0~255 for the display purpose.

Table 1. Comparison of the algorithm-based segmentation and the hand-labeled segmentation on six real TRUS images.

Subjects	Average Distance (Voxels)	Overlap Volume Error (%)	Volume Error (%)
Image1	1.01	3.90	2.06
Image2	0.96	4.04	2.15
Image3	0.95	3.32	1.12
Image4	1.22	4.63	4.47
Image5	1.34	4.87	1.30
Image6	1.13	4.03	2.25
Mean	**1.10**	**4.13**	**2.23**
Stand. Deviation	**0.16**	**0.55**	**1.19**

4 Conclusion

We have proposed an efficient segmentation approach for fast segmentation of prostates from 3D TRUS images. Our segmentation approach was formulated as a multi-resolution framework, and it was speeded up by two techniques, respectively designed for different resolutions. In both low and middle resolutions, Zernike moment-based edge detector is used to replace the step of SVM-based tissue classification and boundary identification, for fast capturing boundary information for deformable segmentation. In the high resolution, a new training method has been designed to increase the efficiency of the finally trained SVM for texture-based tissue classification, thereby equally increasing the efficiency of texture matching step in deformable segmentation procedure. Compared to our previous segmentation method [6], the proposed one is 10 times faster in segmenting 3D prostate from TRUS images, yet without losing any segmentation accuracy.

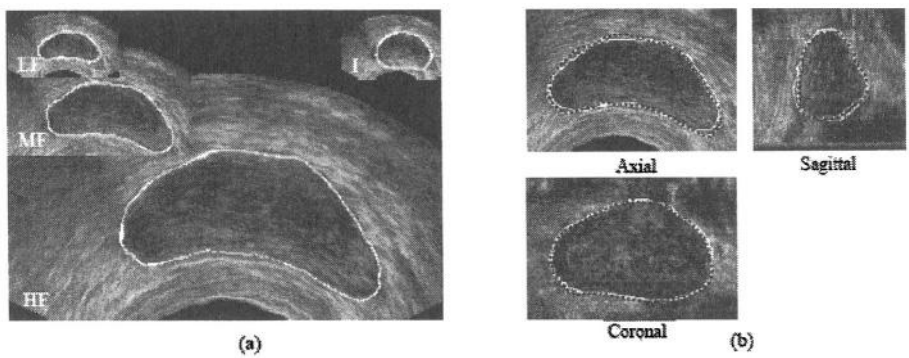


Fig. 5. (a) A typical multi-resolution deformation procedure. The contour denotes the model on a selected slice of the TRUS image. The contour in the image "I" is the initialized model in the low resolution. The contours in the images "LF" "MF" and "HF" denote the finally deformed models in the low, middle and high resolution images. (b) Visual comparisons between algorithm-based and hand-labeled segmentation results. The white contours are the hand-labeled results, while the dashed ones are the algorithm-based segmentation results.

References

1. "Overview: Prostate Cancer", http://www.cancer.org, 2004.
2. Ghanei, H. Soltanian-Zadeh, A. Ratkesicz and F. Yin, "A three-dimensional deformable model for segmentation of human prostate from ultrasound image", Med. Phy., Vol. 28, pp. 2147-2153, 2001.
3. Hu, D. Downey, A. Fenster, and H. Ladak, "Prostate surface segmentation from 3D ultrasound images", ISBI, pp. 613-616, Washington, D.C., 2002.
4. Shao, K.V. Ling and W.S. Ng, "3D Prostate Surface Detection from Ultrasound Images Based on Level Set Method", MICCAI 2002, pp. 389-396, 2002.
5. L. Gong, S.D. Pathak, D.R. Haynor, P.S. Cho and Y. Kim, "Parametric Shape Modeling Using Deformable Superellipses for Prostate Segmentation", TMI, pp. 340-349, Vol. 23, 2004.
6. Y. Zhan and D. Shen, "Automated Segmentation of 3D US Prostate Images Using Statistical Texture-Based Matching Method", MICCAI, 2003, Nov 16-18, Canada.
7. P. Chalermwat and T. El-Ghazaw, "Multi-resolution Image Registration Using Genetics", ICIP, Japan, Oct. 1999.
8. D. Shen and C. Davatzikos, "HAMMER: Hierarchical Attribute Matching Mechanism for Elastic Registration", IEEE Trans. on Medical Imaging, 21(11):1421-1439, Nov 2002.
9. B.S. Manjunath and W.Y. Ma, "Texture Features for Browsing and Retrieval of Image Data", IEEE Trans. on Pattern Anal. Mach. Intell., Vol. 18, pp. 837-842, 1996.
10. S. Ghosal and R. Mehrotra, "Orthogonal Moment Operators for Subpixel Edge Detection", Pattern Recognition, Vol 26, pp. 295-305, 1993.
11. C.J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery, Vol. 2, pp. 121-167, 1998.
12. E. Osuna, F. Girosi, "Reducing the run-time complexity of Support Vector Machines", ICPR, Brisbane, Australia, 1998.

White Matter Lesion Segmentation from Volumetric MR Images

Faguo Yang[1,2], Tianzi Jiang[2], Wanlin Zhu[2], and Frithjof Kruggel[1]

[1] Max-Planck Institute of Cognitive Neuroscience
Stephanstrasse 1A, 04103 Leipzig, Germany
[2] National Lab of Pattern Recognition, Institute of Automation
Chinese Academy of Sciences, Beijing 100080, China.
{fgyang, jiangtz, wlzhu}nlpr.ia.ac.cn

Abstract. White matter lesions are common pathological findings in MR tomograms of elderly subjects. These lesions are typically caused by small vessel diseases (e.g., due to hypertension, diabetes). In this paper, we introduce an automatic algorithm for segmentation of white matter lesions from volumetric MR images. In the literature, there are methods based on multi-channel MR images, which obtain good results. But they assume that the different channel images have same resolution, which is often not available. Although our method is also based on T1 and T2 weighted MR images, we do not assume that they have the same resolution (Generally, the T2 volume has much less slices than the T1 volume). Our method can be summarized as the following three steps: 1) Register the T1 image volume and the T2 image volume to find the T1 slices corresponding to those in the T2 volume; 2) Based on the T1 and T2 image slices, lesions in these slices are segmented; 3) Use deformable models to segment lesion boundaries in those T1 slices, which do not have corresponding T2 slices. Experimental results demonstrate that our algorithm performs well.

1 Introduction

White matter lesions are common pathological findings in MR tomograms of elderly subjects, which are typically caused by small vessel diseases (e.g., due to hypertension, diabetes). It is currently under debate how much the presence of these lesions is related to cognitive deficits in elderly subjects. So an automatic analysis is very useful. But building reliable tools to segment MR images with pathological findings is a nontrivial task. Manual segmentation is a fundamental way to segment MR images, but it takes a trained specialist a lot of time because of the large amount of image data. Moreover, different specialists may give different segmentation results. Compared with manual segmentation, the advantages of automatic segmentation include increased reliability, consistency, and reproducibility.

In the literature, several brain lesion segmentation methods have been introduced [1,2,3,4,5,6,7], and many of them concentrate on multiple sclerosis

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 113–120, 2004.

(MS) [3,4,7]. Some of these algorithms use only T1-weighted images [5,6], others are based on multi-channel volumes [1,3,4]. In [1], a semi-automatic method is introduced, in which typical tissue voxels (white matter, cerebral spinal fluid, and gray matter, lesions) are selected by the user to train an artificial network, then it is used to analyze the MR images; Leemput, et al. [3] view lesions as outliers and use a robust parameter estimation method to detect them. A multi-resolution algorithm is used to detect Multiple Sclerosis lesions in [4]. Kovalev, et al. [5] take advantage of texture analysis to extract features for description of white matter lesions. In [7], models of normal tissue distribution are used for brain lesion segmentation, but they label lesions in the transformed data space, instead of the original image volume.

The main obstacle to white matter lesion segmentation is that the intensities of white matter lesions and gray matter are very similar in T1-weighted images , therefore they can not be distinguished only by intensities of the T1 images (see Fig. 1). It is expected that multi-channel based methods will obtain better results. On most current scanners, it takes an inacceptable long time to acquire a T2-weighted image volume at the same resolution as a T1-weighted volume, that is approximately 1 mm in all spatial directions. Thus, most imaging protocols only allow for the acquisition of a sparse set (20-30) of T2-weighted slices at a typical slice thickness of 5-7 mm. The aim of this paper is to develop a white matter lesion segmentation algorithm based on multi-channel MR volumes, but we do not assume that T1 volumes and T2 volumes have the same resolution. Our algorithm can be summarized as follows: 1) Register the T1 image volume and the T2 image volume to find the T1 slices corresponding to those in the T2 volume; 2) Based on the T1 and T2 image slices, lesions in these slices are segmented; 3) Use deformable models to segment lesion borders in those T1 slices, which do not have corresponding T2 slices. The deformable model is initialized according to the neighboring segmented lesions based on both T1 and T2 slices.

The rest of the paper is organized as follows: Section 2 is devoted to the segmentation of lesions based on both T1 image and T2 image slices. We describe how to apply deformable models for lesion segmentation in Section 3; Experimental results are given in Section 4; A summary is made in Section 5.

2 Lesion Segmentation Based on T1 and T2 Slices

As for the T1 image volume and the T2 image volume are of the same person and are scanned almost in the same time, registration methods based on rigid transformation are enough for our requirements. We use the registration method [8] to find which T1 slices correspond to those T2 slices. And these T1 slices form a T1 volume denoted by S with the same resolution as the T2 volume. At the same time, the T1 volume is transformed using the same transformation parameters.

We firstly segment lesions in those T1 weighted slices that have corresponding T2 slices. These segmented lesions provide some location and shape information

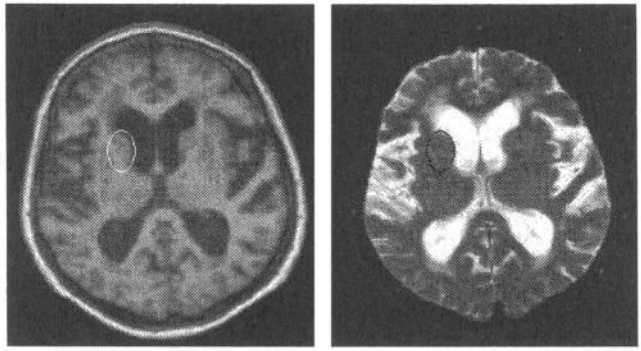


Fig. 1. Illustration of image parts, which can not be distinguished only by T1 image.

of the lesions in other slices. The steps to segment lesions based on T1 and T2 slices are as follows:

- Both the selected T1 image volume S and the T2 volume are segmented using a C-fuzzy mean algorithm [9]. Only those voxels, which are similar with gray matter in T1 channel and similar with CSF in T2 channel are classified as lesions. This can be expressed as follows:

$$\Gamma_{les} = \{v | p_{v,gm}(T1) p_{v,csf}(T2) > \beta\} \tag{1}$$

  where $p_{v,gm}(T1)$ and $p_{v,csf}(T2)$ are the memberships indicating in how much context voxel v belongs to gray matter in T1 volume and belongs to CSF in T2 volume, respectively.
- From the segmented lesions, we can obtain some statistical lesion information (mean value μ_{les} and standard deviation σ_{les}).

Some slices of the T2 image volume, its corresponding T1 slices and the segmented lesions are shown in Fig. 2.

3 Lesion Segmentation by Applying Deformable Models

Following the lesion segmentation in the corresponding slices, it is necessary to process those slices without corresponding T2 slices. We assume that the lesions in neighboring T1 slices are similar, that is, the location and shape of the lesions does not greatly vary. This is likely, when the T1 volume resolution in the slice direction is high (our image data is 1mm) and the lesions are not very small. We can make use of the the location and shape information obtained from the segmented lesions based on both weightings. We use deformable models to accomplish this task. The deformable model is firstly initialized by the neighboring segmented lesions, then adapts itself according to the current image slice.

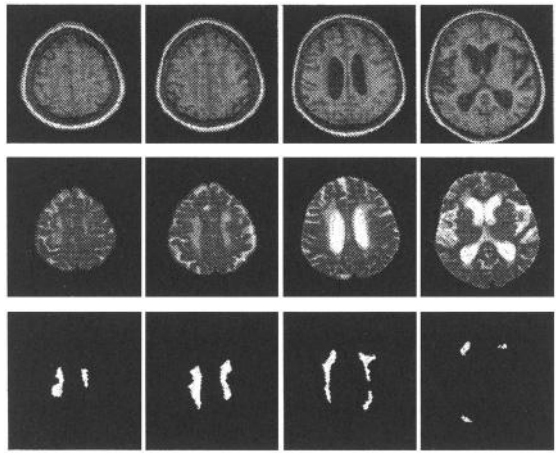


Fig. 2. Original slices and segmented lesions. The first row is the mapped T1 slices; The second row is original T2 slices; And the last row is the segmented lesions.

3.1 Representation of the Discrete Contour Model

In this section, we give a brief description of the original model (refer to [10] for details). The model is made up of a group of vertices connected by edges (see Fig. 3). The position of vertex V_i is represented by vector p_i. The unit vector of the edge between vertex V_i and V_{i+1} is denoted by d_i. The unit tangential vector at vertex i is defined as

$$t_i = \frac{d_{i-1} + d_i}{||d_{i-1} + d_i||} \tag{2}$$

The radial vector r_i is obtained by rotating the tangential vector $\pi/2$ clockwise. Each vertex moves along its radial vector during the deformation. The movement of each vertex is based on the sum of the internal, external, and a damping force. The internal force, which is based on the curvature $c_i = d_i - d_{i-1}$, makes the dynamic contour smooth. The damping force $f_{damp,i}$ is proportional to the

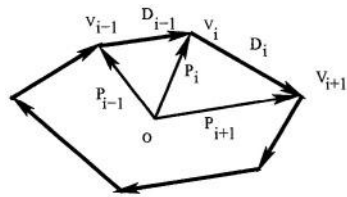


Fig. 3. The model consists of a set of vertices V_i, which are connected by edges D_i.

velocity of the vertex. The contour deformation is computed in discrete positions in time

$$p_i(t+\Delta t) = p_i(t) + v_i(t)\Delta t \tag{3}$$

$$v_i(t+\Delta t) = v_i(t) + a_i(t)\Delta t \tag{4}$$

$$a_i(t+\Delta t) = f_i(t+\Delta t)/m_i \tag{5}$$

$$f_i = w_{ex} f_{ex,i} + w_{in} f_{in,i} + w_{damp} f_{damp,i} \tag{6}$$

where a_i, v_i and m_i are vertex acceleration, velocity and mass, respectively; w_{ex}, w_{in}, and w_{damp} are weights for the external, internal, and damping forces, respectively. For a special application, it is important to define the external force.

3.2 External Forces

For an application of the discrete contour model, it is of great importance to define a proper external force. Our external force includes an edge based component, a balloon component and a shrinking component. Often there are multiple boundaries near the lesions. We assume that the initial contour is placed near the real edge of the lesions. The assumption is true, when the T1 image volume has high resolution along the slice direction and the lesion is not too small. And we also define the balloon and shrinking force to cope with the cases that the assumption is not true. The total external force is:

$$f_{ex,i} = f_{ex,i,ege} + f_{ex,i,bal} + f_{ex,i,shr} \tag{7}$$

- Edge based component: we search the nearest edge in the direction of r_i. The point on this direction can be represented by $p_i + xr_i$. The intensities of the points $p_i + xr_i$ form an function $g(p_i + xr_i)$ with $-L < x < L$, where L is a constant. The nearest edge point is

$$p^\star = p_i + x^\star r_i \tag{8}$$

  where $x^\star$ is the minimum of x, which satisfies $g^{(2)}(p_i + xr_i) = 0$. The edge based component is then calculated by

$$f_{ex,i,ege} = ((p^\star - p_i)\cdot r_i)r_i = x^\star r_i \tag{9}$$

- Balloon component: In some cases, the initial vertex is inside the lesions, so a balloon force is required to let the vertex move outside. We use Γ_i to denote the set made up of vertex i and its four nearest neighbors. if $|I_j - \mu_{les}| < \sigma_{les}$, $j \in \Gamma_i$, then $f_{ex,i,bal} = -\gamma$. Here I_j is the intensity of pixel j; μ_{les} and σ_{les} are the mean value and standard deviation of the lesions, respectively, which are obtained during the lesion segmentation based on both weightings.
- Shrinking component: The initial vertex sometimes locates outside the lesions, therefore we add a shrinking component to let the vertex move inside. if $|I_j - \mu_{les}| > 2\sigma_{les}$, $j \in \Gamma_i$, then $f_{ex,i,shr} = \gamma$.

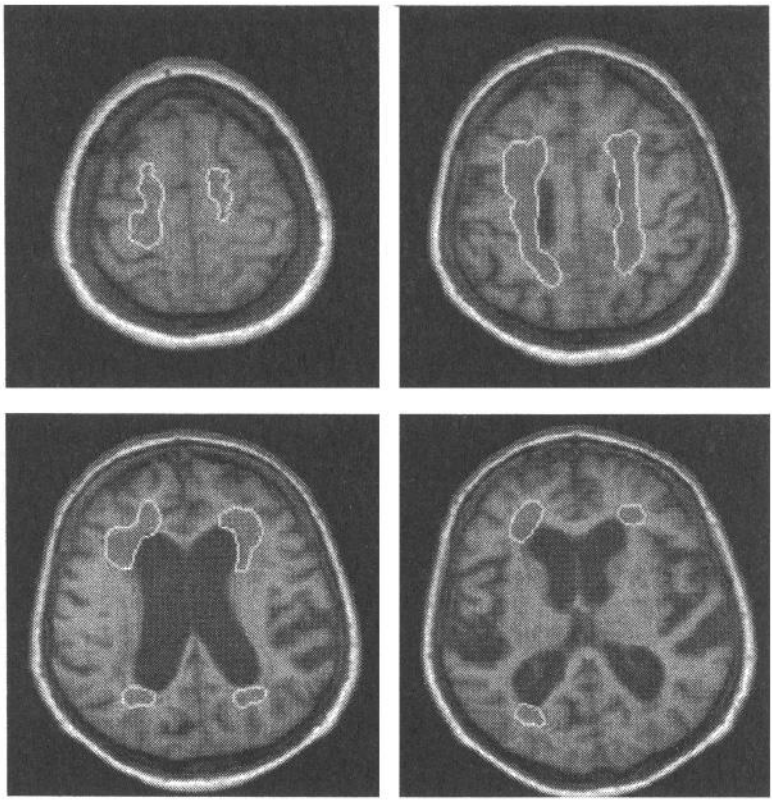


Fig. 4. Some slices of segmented lesions of one patient.

4 Experimental Results

Our experimental image volumes were obtained by a 1.5 Tesla clinical MR scanner. The voxel size of the T1 volume is $0.977 \times 0.977 \times 1$ mm; The T2 volume is $0.498 \times 0.498 \times (5.0 - 7.0)$ mm.

In our experiment, some important parameters are as follows: $\beta = 0.2$, $\Delta = 0.5$, $m_i = 1$, $w_{ex} = 2.5$, $w_{in} = 2$, $w_{damp} = 1$, $L = 5$, $\gamma = 1$. In Fig. 4 and Fig. 5, some slices of the segment results of different patients are displayed, in which the boundaries of the lesions are shown. Validating the extent of the white matter lesions is difficult: Lesion borders are faint, and sometimes the distinction between a lesion and grey matter of a fundus region is hard to draw. Thus, we resort to a critical visual inspection of the results by a neuroradiologist. Note that the caudate nucleus that is similar in intensity to grey matter and is often adjacent to a white matter lesion, is correctly excluded from the lesion area here.

The effect of the internal force is to make the curve smooth. The larger the parameter w_{in}, the smoother the curve. The external force try to let the curve approach the image edges. The final obtained curve is a trade-off between the

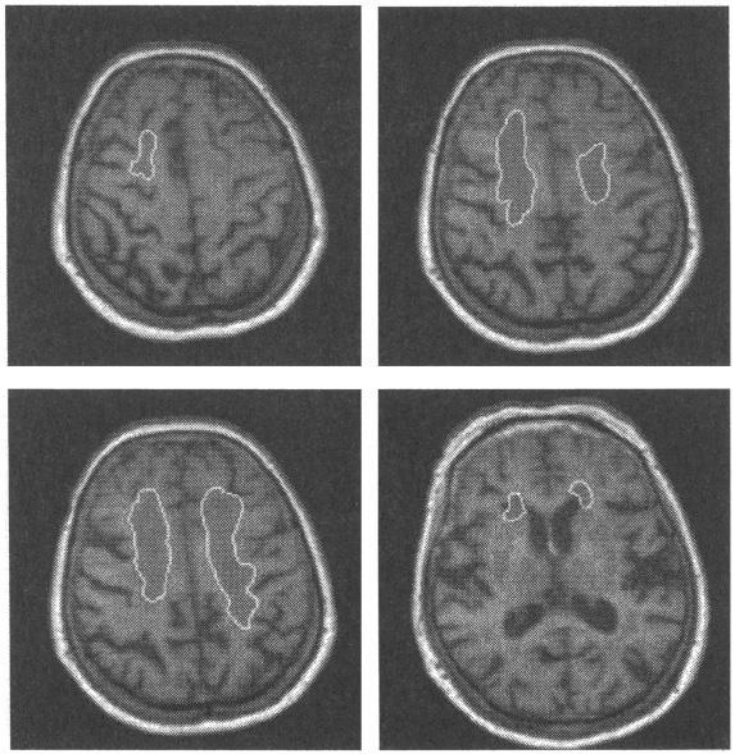


Fig. 5. Some slices of segmented lesions of another patient.

external force and internal force. For our image datasets, the above parameter values are proper. It seems that they are too many parameters. In fact, for a set of image volumes, we can adjust them once and they can be used for all other image volumes. In all of the parameters, the user must adjust β. For our image volumes, $\beta = 0.1 - 0.2$ is proper, which can be seen from Fig. 6. Because the deformable model adapts itself based on the image slices, the result is not sensitive to parameter β.

5 Summary

In this paper, we developed a novel and effective white matter lesion segmentation algorithm. Our method is based on T1 and T2 image volumes. But we do not assume that they have the same resolution. We firstly analyze those T1 slices, which have corresponding T2 slices. The segmented lesions in these slices provide location, shape and intensity statistical information for processing other neighboring T1 slices without corresponding T2 slices. This prior information is used to initialize a discrete contour model in the segmentation of the remaining T1-weighted slices.

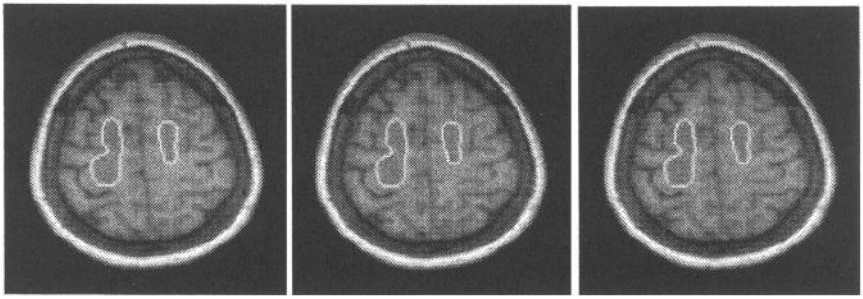


Fig. 6. Effect of parameter β, from left to right $\beta = 0.1, 0.15,$ and 0.2, respectively.

References

1. A. P. Zijdenbos, B. M. Dawant, R. A. Margolin, and A. C. Palmer, "Morphometric Analysis of White Matter Lesions in MR Images: Method and Validation", IEEE Trans. Med. Imag., vol. 13, 1994, pp. 716-724.
2. I. Kapouleas, "Automatic Detection of White Matter Lesions in Magnetic Resonance Brain Images", Comput. Methods and programs in Biomed., vol. 32, 1990, pp. 17-35.
3. K. V. Leemput, F. Maes, D. Vandermeulen, A. Colchester, and P. Suetens, "Automated Segmentation of Multiple Sclerosis Lesions by Model Outlier Detection", IEEE Trans. Med. Imag., vol. 20, 2001, pp. 677-688.
4. C. Pachai, Y. M. Zhu, J. Grimaud, M. Hermier, A. Dromigny-Badin, A. Boudraa, G. Gimenez, C. Confavreux, J. C. Froment, "A Pyramidal Approach for Automatic Segmentation of Multiple Sclerosis Lesions in Brain MRI", Computerized Medical Imaging and Graphics, vol. 22, 1998, pp. 399-408.
5. V. A. Kovalev, F. Kruggel, H. J. Gertz, D. Y. V. Cramon, "Three-Dimensional Texture Analysis of MRI Brain Datasets", IEEE Trans. Med. Imag. vol. 20, 2001, pp. 424-433.
6. S. A. Hojjatoleslami, F. Kruggel, and D. Y. von Cramon, "Segmentation of White Matter Lesions from Volumetric MR images", Medical Image Computing and Computer-Assisted Intervention, Lecture Notes in Computer Science, vol. 1679, pp. 52-61, Heidelberg Springer.
7. M. Kamber, R. Shinghal, D. L. Collins, G. S. Francis, and A. C. Evans, "Model-Based 3-D Segmentation of Multiple Sclerosis Lesions in Magnetic Resonance Brain Images", IEEE Trans. Med. Imag., vol. 14, 1995, pp. 442-453.
8. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality Image Registration by Maximization of Mutual Information", IEEE Trans. Med. Imag. vol. 16, 1997, pp. 187-198.
9. M. N. Ahmed, S. M. Yamany, N. Mohamed, A. A. Farag and T. Moriarty, "A Modified Fuzzy C-Means Algorithm for Bias Field Estimation and Segmentation of MRI Data", IEEE Trans. Med. Imag., vol. 21, 2002, pp. 193-199.
10. S. Lobregt, and M. A. Viergever,"A Discrete Dynamic Contour Model", IEEE Trans. Med. Imag., vol. 14, 1995, pp. 12-24.

Active Shape Model Segmentation Using Local Edge Structures and AdaBoost

Shuyu Li, Litao Zhu, Tianzi Jiang

National Laboratory of Pattern Recognition, Institute of Automation,
Chinese Academy of Sciences, Beijing, People's Republic of China
{syli, ltzhu, jiangtz }@nlpr.ia.ac.cn
http://nlpr-web.ia.ac.cn/English/mic

Abstract. The paper describes a machine learning approach for improving active shape model segmentation, which can achieve high detection rates. Rather than represent the image structure using intensity gradients, we extract local edge features for each landmark using steerable filters. A machine learning algorithm based on AdaBoost selects a small number of critical features from a large set and yields extremely efficient classifiers. These non-linear classifiers are used, instead of the linear Mahalanobis distance, to find optimal displacements by searching along the direction perpendicular to each landmark. These features give more accurate and reliable matching between model and new images than modeling image intensity alone. Experimental results demonstrated the ability of this improved method to accurately locate edge features.

1 Introduction

Segmentation is one of important areas in image analysis. Previous edge-based method focus on the analysis and design of filters for detection of local image structure, such as edges, ridges, corners and T-junctions. Among them, steerable filters [1][2] can provide any orientation information, which are sensitive to the presence of edges, bars and other simple image structure. Unfortunately, it is difficulty to obtain good results only to use this kind of edge-based method under poor conditions.

Active contours or snakes [3][4], and level sets [5] can usually deform freely to reach the real boundaries of objects by energy minimization, given any initial curve (surface) and regardless of topological variations. But such methods have little priori knowledge incorporated, and the assumption about object boundaries coinciding with edges or ridges is usually not tenable.

Active shape models (ASM) proposed by Cootes et al. [6], provide popular shape and appearance models for object localization. The method makes full use of priori shape and appearance knowledge of object and has the ability to deform within some constraints. Based on modeling local features accurately, ASM obtains good results in shape localization [9][10].

In recent years, some segmentation approaches based on ASM are found in the literature. They mainly focus on two fold: 1) image description; Rather than intensity gradients, several image features are extracted to represent local image structure. In

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 121-128, 2004.

order to implement accurate and reliable matching between models and new images, I.M. Scott et al. [11] used gradient orientation, corner and edge strength as local structure descriptors. Bram van Ginneken et al.[7] used the moments of local histograms extracted from filtered versions of the images by a filter bank of Gaussian derivatives to construct local image structure descriptors. Jiao et al. [12] used Gabor-Wavelet features to model local structures of the image. 2) Modeling local features; One simple way to model features is to calculate the mean jet of each landmark in all training images. Then for every landmark, the PCA is applied to model the variability of local features descriptors. When feature space is very high or the distribution of samples is non-Gaussian, the above modeling method doesn't work well. Bram van Ginneken et al. [7] provided a new approach to model local structure features. They perform a statistical analysis to learn which descriptors are the most informative at each resolution, and at each landmark. The kNN classifier with the selected set of features by sequential feature forward and backward selection is constructed at each landmark. Jiao et al. [14] used an E-M algorithm to model the Gabor feature distribution.

There are two contributions of our work. One contribution is to extract local edge orientation features using steerable filters as local image structure descriptions. This, to a large extent, compensates deficiency in the search based on intensity gradients. The other contribution is to introduce a machine learning algorithm to model matching, which construct a classifier for each landmark by selecting a small number of important features using AdaBoost [8].

The rest of the paper is organized as follows: The original ASM algorithm is briefly described in Section 2. In Section 3, we present our edge-based representation for local structure of shape, and a machine learning algorithm based on AdaBoost is used to construct efficient classifier for local edge structure, then how to search is presented. Experimental results are presented in Section 4, and conclusions are drawn in Section 5.

2 Active Shape Model

2.1 Shape Model

In order to capture the variability in shape from training examples, a 2-D point distribution model can be built. Each training image is represented by n manual landmark points. Each point represents a particular part of the object or its boundary, so must be placed in the same way on every training example. The landmarks from each image are represented as a vector x and then aligned by Procrustes Analysis [15].

$$X = (x_1, y_1, \ldots, x_n, y_n)^T \tag{1}$$

The modes of variation can be found by applying principle component analysis [13] to the deviations from the mean.

Any shape in the training set can be approximated using the mean shape and a weighted sum of these deviations obtained from the first t modes

$$X = \overline{X} + Pb \tag{2}$$

where b is a vector of t shape parameters. Different b corresponds to different shape in the model. By constraining the value of parameter b, we can ensure the generated shape is in the range of the allowable variation of the shape model.

2.2 Local Appearance Models

In order to guide the matching between the model and the object, it is necessary to construct local appearance model for each landmark in the training examples.

In ASM, the local image feature of each landmark is represented by sampling intensity gradients along the profile perpendicular to the landmark contour. It is assumed that these local image features are distributed as a multivariate Gaussian for each landmark. Then the similar appearance model can be constructed by deriving the mean profile and the covariance matrix from the profile examples. The matching between the current positions in test image to the corresponding model is determined by minimizing the Mahalanobis distance from the feature vector of the landmark to the corresponding model mean.

The ASM search procedure is implemented by using the local appearance model to find a new shape and then updating the model parameters to best fit the new search shape on each iteration. In order to fast convergence, multiresolution framework is adopted. The model does match to the object in the way from the coarse to fine resolutions.

3 ASM with Local Edge Structure and AdaBoost

In this section, a new appearance model is described that use steerable filters to extract local image edge features, then a machine learning algorithm based on AdaBoost is used to construct effective classifiers for each landmark. The aim is to move the landmark points to better locations during optimization along a profile perpendicular to the object contour. The best location is that its local edge structure is the most similar to corresponding landmark. First, a point distribution model is also constructed as the same as the original ASM.

3.1 Local Edge Structure Detector

3.1.1 Local Edge Features

In this paper, we have used steerable filters to extract local image edge features. Steerable filter is to describe a class of filters in which a filter of arbitrary orientation is synthesized as a linear combination of a set of basis filters. Steerable filters [1][2], are excellent for the detailed analysis of boundaries, which are sensitive to the presence of edges, bars and other simple image structure. Steerable filters can provide any

orientations information. This, to a large extent, compensates deficiency in the search based on intensity gradients.

Assume the $G(x, y; \sigma)$ is Gaussian kernel function, then $G_1^{0°}$ is the first x derivative and $G_1^{90°}$ is the first y derivative of Gaussian function.

A G_1 filter at an arbitrary orientation θ can be synthesized by taking a linear combination of $G_1^{0°}$ and $G_1^{90°}$:

$$G_1^{\theta} = \cos\theta G_1^{0°} + \sin\theta G_1^{90°} \tag{3}$$

Similarly, a G_2 filter at an arbitrary orientation θ can be represented by the linear combination of $G_2^{0°}$, $G_2^{90°}$ and $G_2^{0°90°}$.

$$G_2^{\theta} = \cos^2\theta G_2^{0°} + \sin^2\theta G_2^{90°} - 2\cos\theta\sin\theta G_2^{0°90°} \tag{4}$$

The derivatives of Gausssian filters offer a simple illustration of steerability. After the original image convoluting with different Gaussian filters, different filtered versions can be obtained.

For each landmark, there are different features when varying some parameters: the order of Gaussian derivatives, the number of Gaussian kernel scales σ, and the angles θ. The responses of different orientation Gaussian filters can be obtained by varying the angles θ. In this paper, we use the zero, first and second-order derivatives (G, G_1 and G_2), four inner scales (σ=0.5, 1, 2, 4 pixels) and sixteen orientation (θ =0, π/8, π/4, …15π/8). The total number of features of each point is (l+16×2)×4=132. Obviously, different features can be extracted by using more scales, orientations and higher-order derivatives.

In order to describe the local edge structure, we use a square grid of Ngrid×Ngrid (Ngrid=7) points and the landmark point at the center of the grid representing the structure of each landmark. As each point within the grid was extracted 132 features, then for each landmark, 7×7×132=6468 features were used to represent the local edge structure. Comparing with the original ASM, these features reflect more subtler local image structure.

3.1.2 Modeling Local Structure Using AdaBoost

For a given point, 6468 features were computed to represent the local edge structure. Given a feature set and a training set of edge points and non-edge points, any machine learning approaches could be used to learn a classification function. In this paper, AdaBoost is used both to select a small set of critical features and train the effective classifiers [8]. The AdaBoost learning algorithm is one of the integrated machine learning algorithms, its main virtue is to combine some weak classifiers (the correct rate larger than 0.5) to a strong classifier. As it is very easy to train weak classifiers, the AdaBoost learning algorithm can be applied widely. More importantly, a number of results were later proved that AdaBoost algorithm has good generalization ability [14]. The weak learning algorithm is designed to select the single local edge feature which best separates the specialized edge points and non-edge points. For each feature, the weak learning classifier determines the optimal threshold classification function, so that the minimum number of examples is misclassified.

Each selected feature corresponds to a weak classifier. After T iterations, T weak classifiers $h_j(x)$ were combined to a final strong classifier. The detailed AdaBoost algorithm is referred in the literature [8].

3.2 Model Searching Procedure

When the model is fitted to a test image, the scheme starts by computing the 6468 features for each searching point. Instead of sampling the normalized derivative profiles, the feature set at each position along the profile perpendicular to the object contour is fed into a trained classifier to determine the position that this pixel is moved to. The output 1 represents that the probability that the pixel is on the edge is large, 0 means that the probability that the pixel is not on the edge is large. The index along the profile is oriented from non-edge position to the edge position.

4 Experimental Results

For experiment, a collection of 36 slices of the brain was used, in which the corpus callosum were labeled by manual. The resolution is 256×112 pixels. On each image 36 landmarks are labeled. Since we did not have a large data set, we performed leave-1-out experiments, by repeatedly training an ASM on 35 of the images and testing it on the remaining image.

For each parameter of ASMs, a fixed setting was selected that yielded good performance, after initial pilot experiments. The other settings were three levels of resolution, ten iterations/level, profiles of length seven and evaluation of eleven positions/iteration for ASM method. For the extended ASMs, we adopt single resolution method. At most two hundreds features were selected from 6468 features for each landmark. Training data were selected from 7×7 neighborhoods around each landmark (Ngrid=7). In the AdaBoost algorithm, the weak classifier is linear discriminative classifier. One of search results is showed in Figure 1.

After the ASM search had converged, in order to evaluate the search result, we measured the distance from each search shape and the manually labeled shapes in two ways. One way is that we calculate the displacement between each searched point and the corresponding labeled point. The distribution is shown in Figure 2a. The x-coordinate is the displacement (in pixels) between the searched points and labeled point locations. The y-coordinate is the percentage of points whose displacement to the target is x. From this figure, we can see that the extended ASM achieves more accurate results than ASM.

The other way is that we calculate the overall displacement of the search shape to the labeled shape for each test image. The distance of two shapes is defined as the sum of all distance between corresponding points. We calculate DisA (the distance between ASM search shapes and the labeled shapes) and DisAA (the distance of ASM-AdaBoost search shapes to the labeled shapes). Then we calculate the difference

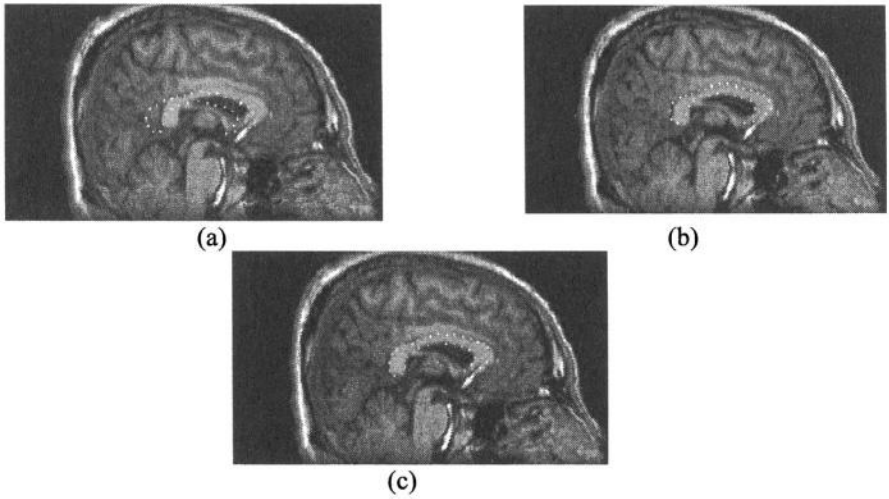


Figure 1 Example result for segmenting the corpus callosum (a) the initial shape (b) the segmentation shape by ASM (c) the segmentation result by the improved ASM method

$$m = (DisA - DisAA)/DisA \times 100\% \quad (5)$$

which measures the percentage of improvement of DisAA. In Figure 2b, the x-coordinate is the index of test images, and the y-coordinate is its corresponding m value. The blue point showed the m>0, which means the result of ASM-AdaBoost is better than ASM. From this figure, we can see that ASM-AdaBoost works worse than in 6 test images, and works better than ASM in the remaining 30 test images.

Our algorithm is tested on a PIV 2.8G computer with 512M memory. The program is in Matlab. The training process is off-line. For each training images, it cost about 2.5 minutes for training one classifier. The feature images have to be computed on-line (during segmentation), which required 0.8s each iteration. The total time for segmenting one image was about 20s for the original ASM scheme and 390s for the extended method.

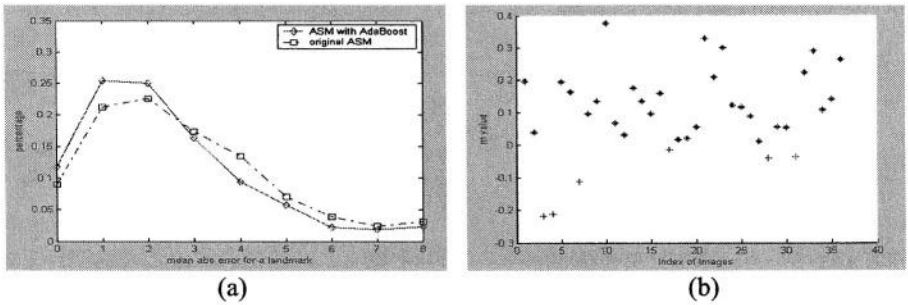


Figure 2 (a) Point-to-point displacement (b) Shape-to-shape distance difference

5 Discussion

In this section, we discuss some properties and refinements of the proposed ASM method and future work.

The main advantage in this paper is that the information on the edge structure is used. For each landmark, a specialized classifier is designed to assign each image location to one of two classes: edge or non-edge by machine learning algorithm. We have conducted experiment that indicated the performance of such an approach is better than the original ASM. The set of selected features varies considerably/landmark and would be different for different application. We used AdaBoost algorithm to select critical features and construct the effective classifiers.

In the process of training, how to define the training samples is very important. One can classify the outside/inside edge pixels. But for the ridge, it is difficult to determine which position is outside or inside the edge. So we represent the image structure using local edge structure. Experiments show the representation can obtain good segmentation results. But from Figure 2b, we find that the segmentation results of the 3th and 4th images by the extended ASM method are obviously worse than those of original ASM. At the same time, we find this two images have larger rotation than other images. However, the extracted features have relation to orientation. We consider that good results can be obtained if extracting features and training classifiers after all training images are aligned. But in our experiment, extracting feature has conduct without aligning all training images. We also extract rotation-invariant features for discriminating edges.

We conclude that the new extended ASM method improves the original method through the use of machine learning algorithm to model local edge structure well.

References

1. W. T. Freeman, E. H. Adelson. The design and use of steerable filters, IEEE Transaction on Pattern Analysis and Machine Intelligence, 13(1991) : 891-906
2. H. Greenspan, S. Belongie, R. Goodman, P. Perona, S. Rakshit, C. Anderson, Overcomeplete steerable pyramid filters and rotation invariance. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (1994)
3. M. Kass, A. Witkin, D. Terzopoulos, Snakes: Active contour models, Int. J. Comput. Vis., 1(1988)321-331
4. T. McInerney, D. Terzopoulos, Deformable models in medical image analysis: A survey, Med. Image Anal., 1(1996), 91-108
5. J. A. Sethian, Level set methods and fast marching methods, 2ed. Cambridge, U.K.: Cambridge Univ. Press, (1999)
6. T. F. Cootes, C. J. Taylor, D. H. Cooper, J. Graham, Active shape models- their training and application, Computer vision and image understanding, 61 (1995), 38-59
7. B. van Ginneken, A. F. Frangi, Joes J. Staal, Bart M. ter Haar Romeny, Max A. Viergever, Active Shape Model Segmentation With Optimal Features, IEEE Trans. On Medical Imaging, 21(2002), 924-933
8. Y. Freund, R. E. Schapire, A decision-theoretic generalization of on-line learing and an application to boosting, In Computational Learning Theory: Eurocolt'95, Springer-Verlag,(1995) 23-37

9. T. F. Cootes, A. Hill, C. J. Taylor, J. Haslam, The use of active shape models for locating structures in medical images, Image and Vision Computing, 12 (1994), 355-365
10. S. Solloway, C. J. Taylor, C. E. Hutchinson, J. C. Waterton. Quantification of articular cartilage from MR images using active shape models. In 4th European Conference on Computer Vision, Cambridge, England, (1996) 400-412
11. I.M.Scott, T.F.Cootes, C.J.Taylor, Improving Appearance Model Matching Using Local Image Structure, Proc. Information Processing in Medical Imaging (2003), 258-269
12. F. Jiao, Stan Z. Li, H. Y. Shum, D. Schuurmans. Face Alignment Using Statistical Models and Wavelet Features, Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, 1(2003), 321-327
13. R.A. Johnson, D.W. Wichern: Multivariate Statistics, a Practical Approach, Chapman & Hall, (1988)
14. R. E. Schapire, Y. Freund, P. Bartlett, W. S. Lee. Boosting the margin: A new explanation for the effectiveness of voting methods. In Proceedings of the Fourteenth International Conference on Machine Learning, (1997)
15. C. Goodall, Procrsutes methods in the statistical analsis of shapes, J. Roy. Statist. Soc. B, 53(1991), 285-339

Segmental Active Contour Model Integrating Region Information for Medical Image Segmentation*

Xin Ran and Feihu Qi

Department of computer science and engineering,
Shanghai Jiaotong University, Shanghai
simonran@sjtu.edu.cn

Abstract. A segmental active contour model integrating region information is proposed. Different deformation schemes are used at two stages for segmenting the object correctly in image plain. At the first stage the contour of the model is divided into several segments hierarchically that deform respectively using affine transformation. After the contour is deformed to the approximate boundary of object, a fine match mechanism using statistical information of local region is adopted to make the contour fit the object's boundary exactly. The experimental results indicate that the proposed model is robust to local minima and able to search for concave objects.

1 Introduction

Active contour models [1], also known as snakes, have been used extensively in image analysis, especially in medical or biological imaging applications. Current active contour model research can be broadly classified into three main branches: a) research on image forces (e.g. [2]); b) research on internal forces (e.g. [3]); and c) research on curve representation (e.g. [4,5]). However, most of those reformative methods usually only incorporate edge information, possibly combined with some prior expectation of shape [6,7]. Ronfard[8] introduced a contrast measure based on a region statistical image model, but there are still many problems when using region-based methods to deform the contour. For example, the efficient control in the course of optimization is difficult and the convergence is very slow unless the initial contour is close to the desired image features.

In this paper a segmental active contour model integrating region information is proposed. The motivation for our work is to present a reformative active contour model combining edge-based method with region-based method to improve the efficiency of traditional models. Different deformation strategies are used separately at two stages for segmenting the object correctly in image plain. At the first stage the contour of the model is divided into several segments hierarchically that deform respectively using affine transformation and gradient-based

* The project supported by The National Natural Science Foundation of China (No.60271033)

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 129–136, 2004.

edge detection technique. Then, a fine match mechanism is adopted to make the contour fit the object's boundary exactly at the second stage, which uses statistical information of local region to redefining the external energy of the model.

2 Segmental Active Contour Model

Segmental active contour model extends the basic concept of active contour model, which improve the computation of internal energy and external energy. The main characteristic of the model is hierarchical deformation mechanism. The purpose of this deformation process is to obtain the approximate boundary of object, and the resulting contour will be processed at next stage.

2.1 Energy Definition of Segmental Active Contour Model

Let's define the contour of model by an ordered set of control points,$\{V_i = (x_i, y_i), i = 1, 2, , N\}$. As customary, the total energy of segmental active contour model is defined as the weighted summation of several energy terms:

$$E_{model} = \omega_1 E_{continue} + \omega_2 E_{smooth} + \omega_3 E_{image} \tag{1}$$

where $E_{continue}$ and E_{smooth} are internal energy of model, which enforce connectivity and smoothness along the contour of model. E_{image} is external energy derived from the image data. ω_1, ω_2, ω_3 are the normalized parameters of the energy terms respectively. The definitions of these energy terms will be described as follows. The continuity term in internal energy function is defined as:

$$E_{continue} = \sum_{i=1}^{N} (d - \|V_i - V_{i-1}\|)^2 \tag{2}$$

where d is the average distance between the control points of the segmental active contour model. $E_{continue}$ attempts to keep the control points at equal distances, which can keep the continuity of the contour and prevent the contour from collapsing or leaking out at the gaps of the object's boundary.

For computing the smoothness term, we use a simple and computationally efficient method, which is proposed by Doug P.Perrin[3]. The smoothness term is defined as:

$$E_{smooth} = \sum_{i=1}^{N} \left\| \theta(V_i) - \frac{\theta(V_{i-1}) + \theta(V_i) + \theta(V_{i+1})}{3} \right\|^2 \tag{3}$$

where $\theta(V_i)$ is the exterior angle of the control point V_i that is formed by the extended line $V_{i-1}V_i$ (dashed line) and line V_iV_{i+1}, as shown in Fig.1. When V_i is updated to its new position, its exterior angle should be one third of the summed exterior angles at control points V_{i-1}, V_i and V_{i+1}. This will produce a

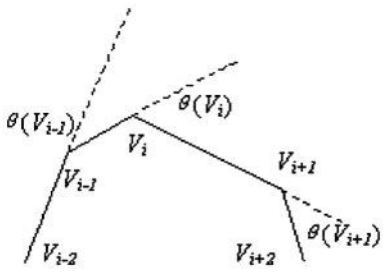


Fig. 1. The illustration of external angle of V_i

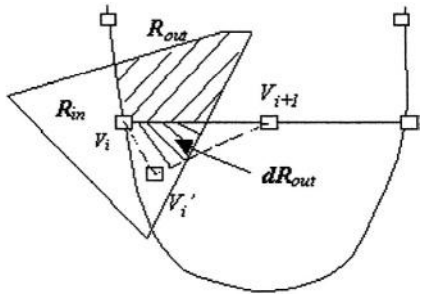


Fig. 2. Illustration of convergence into boundary concavity

constant third derivative over the contour segment from V_{i-2} to V_{i+2}, provided that the segment lengths are equal.

The external force is generally defined to be the negative gradient of a potential function. However, both the magnitude and the direction information of the image gradient should be considered carefully, especially in the case that the object has weak edge or the image has a lot of noise. For each control point, the external energy is defined as follows:

$$E_{image} = \sum_{i=1}^{N} (1 - |\nabla I(V_i)| \cdot |n(V_i) \cdot h(V_i)|) \tag{4}$$

where $|\nabla I(V_i)|$ is normalized magnitude of the gradient of the control point V_i; $h(V_i)$ is the direction of the gradient. $n(V_i)$ is the normal vector of the contour at V_i, which directs towards the snake interior. Accordingly, the external energy will be small where the magnitude of the gradient is high and the direction of image gradient is similar to the normal vector of the contour.

2.2 Hierarchical Deformation Mechanism with Affine Transformation

The deformation process of segmental active contour model is practically the minimizing process of the energy function, hierarchically from global to local, and from coarse to fine. The model contour is divided into several segments. The middle control point along the segment can be defined as the driving point, and each of segments is controlled by its driving point. Suppose that control point V_r is the selected driving point of the rth segment, V_{r-s} and V_{r+s} are control points at the end of the rth segment. If V_r is moved to the position V_r' under the influence of internal force and external force, which can be regarded as an affine transformation in 2-d plane and defined as follows:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad \text{or} \quad \mathbf{X'} = \mathbf{AX} \tag{5}$$

The three points V_r', V_{r-s} and V_{r+s} together with the points V_r, V_{r-s} and V_{r+s} that are not collinear can confirm an affine transformation **A** uniquely. Consequently, the other control points in the rth segment can move to the new positions by the same **A.** This method is robust to local minima since it doesn't deform the control points individually, but it deform the segments of model at a time.

The number of the segments of model is not changeless. At the initial steps, the contour is divided into a small number of segments, thus the number of the driving points that are considered is small and the search area is relatively larger. As the iteration process continues, the number of the segments increases, so more and more control points are selected to be the driving points. The search area becomes smaller such that the shape of the model is smoothly changed. The hierarchical deformation mechanism can be summarized as follows:

1. Initialize the segmental active contour model, including dividing the contour into two segments, selecting the driving points, and computing the gradient field of image and initial energy of model.

2. For every selected segment that has three or more control points, seek the appropriate position of driving point in its searching field and determine the best affine transformation configuration by minimizing the model's energy, then use (5) to move the other control points towards their new positions by the same affine transformation.

3. Divide each segment that has three or more control points into two subsegments, select the driving point if subsegment has more than three control points, compute the internal energy and external energy of model. If no segment can be divided or energy difference between this iteration and the last iteration is less than the threshold E_t, the algorithm is finished, otherwise, go to step 2.

3 Fine Match Strategy Based on Region Information

Traditional active contour models usually use edge-detection method based on gradient to find object's boundaries. However, inside the object's region, both derivatives of either the image or the gradient function become very small or even vanish, therefore providing no clue to the energy-minimizing process. The region-based method introduced by [8] adopted heuristic optimization technique and diffusion processes to obtain a robust active contour modelling scheme, but the convergence is difficult and time-consuming in case that the initial contour of model is not close to object's boundary. In the rest of this section a more effective method is proposed, which makes use of the result of the first stage described in previous section.

3.1 Redefining the External Energy Based on Region Information

The external energy of model is redefined based on statistical information of region around contour neighborhood. The basic idea is that a given closed contour

$\mathbf{C}$ partitions the image plane into an inside(or object) region $\mathbf{R}_{in}$ and an outside (or background) region $\mathbf{R}_{out}$, which have different mean intensity and mean square error(MSE). Suppose that the gray-scale value of image pixel is $I(x,y)$, the mean intensity of region R is I_R, then MSE of intensity in this region can be defined as follows:

$$\delta_R = \frac{1}{n}\sum_R (I(x,y))^2 - I_R^2 \tag{6}$$

where n is the number of pixel in region R. Region R is divided into inside region R_{in} and outside region R_{in} by contour, and the similarity of them can be defined as:

$$D_R[R_{in}, R_{out}] = \frac{1}{|\delta_R - \delta_{R_{in}}| + |\delta_R - \delta_{R_{out}}|} \tag{7}$$

If the contour of model is localized on the true boundary of object, the MSE of region R_{in} and R_{out} will be minimized, and therefore the similarity of the two regions will reach minimal value. Thus, the external energy of segmental active contour model can be redefined using similarity of region:

$$E_{image} = \omega_R \sum_{i=1}^{n} D_i[R_{in}, R_{out}] \tag{8}$$

where ω_R is the normalized parameter of the external energy, $D_i[R_{in}, R_{out}]$ is the similarity of R_{in} and R_{out} in search area of control point V_i.

3.2 Fine Match Scheme and Algorithm

Using segmental active contour model, the contour is already close to the object boundary, therefore fine match scheme only need to search the neighborhood around the contour to exactly localize the object. At each iteration, the control points along the contour of model seek its search area and move to new place to minimize the energy of model.

The search area of control point V_i is the region specified according to the position of V_i and its neighboring points, in which V_i can move at arbitrary direction to minimize the model energy. Both the width and depth of search area should be considered. Connect V_i and its neighboring control points V_{i-1} and V_{i+1}, define the middle point of these two line segments as V'_{i-1} and V'_{i+1}, and compute their normal vectors respectively by :

$$n(V'_{i-1}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \cdot \left(\frac{V'_{i-1} - V_{i-1}}{\|V'_{i-1} - V_{i-1}\|} + \frac{V_i - V'_{i-1}}{\|V_i - V'_{i-1}\|} \right) \tag{9}$$

$$n(V'_{i+1}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \cdot \left(\frac{V'_{i+1} - V_i}{\|V'_{i+1} - V_i\|} + \frac{V_{i+1} - V'_i}{\|V_{i+1} - V'_i\|} \right) \tag{10}$$

Thus, the left and right edge of search area are confirmed by the normal vectors of V'_{i-1} and V'_{i+1}. If choice appropriate depth of search area for a closed contour,

the width definition approach will form a ring-shaped search band around the contour of model, which not only avoids the self-intersection but also guarantees the integrality of search area for whole contour.

A fast adaptive algorithm is proposed for determining the depth of search area. Conforming to the direction of normal vector of the control point, The depth of search area extends outside and inside the model so that the model can expand or contract. Because of the pre-processing by segmental active contour model at previous stage, the initial depth of search area can be set relatively small. If the control point V_i can move to a new place to make the similarity between R_{in} and R_{out} smaller at this iteration, its depth of search area remains unchanged at next iteration. Otherwise, the depth of search area may be reduced. This adaptive method makes minimizing process focus on the deformed part of model and results in faster convergence.

In case of searching boundary concavity, edge-based segmentation methods can not perform well unless using some special process like the method proposed in [2] to diffuse the external force. The fine match scheme presented in this paper is based on statistical information of search areas and more effective to pull the contour into boundary concavity, which is illustrated in Fig.2. The search area of V_i may be divided into two parts(R_{in} and R_{out}, as shown with shadow in Fig.2) by the line segments connected V_i with its neighboring control points. Suppose that V_i is moved to V_i', the search area is then divided into two new part by three line segments connected V_i, V_i' and their neighboring control points. That is to say, the area dR_{out} that belonged to R_{in} before belongs to R_{out} now. Accordingly, the MSE of intensity in either R_{in} or R_{out} is reduced and the similarity between R_{in} and R_{out} becomes smaller. In this way the contour may move into boundary concavity increasingly within finite iteration steps.

The algorithm of fine match scheme using statistical information is summarized as follows:

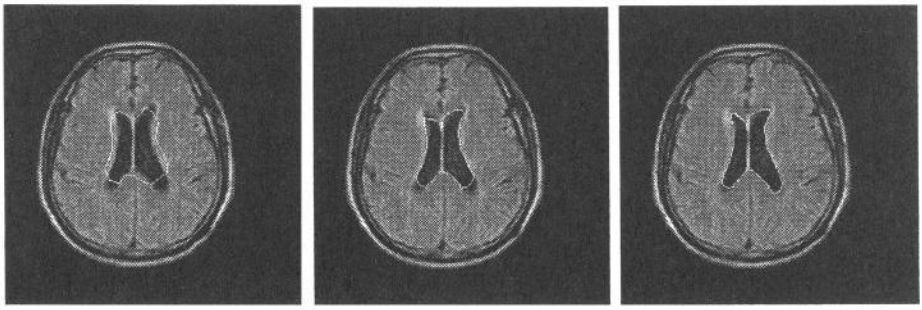
1. Determine the width of search area for each control point using (9) and (10), and specify the initial depth of search area. Compute the similarity between R_{in} and R_{out} in each search area and the initial energy of model.

2. Seek the appropriate position for each control point in its search area to minimizing the model's energy. If the similarity between R_{in} and R_{out} in search area of V_i is reduced less than the threshold D_t, then do not move V_i and reduce its depth of search area.

3. If the energy difference of model between this iteration and the last iteration is less than the threshold E_t, the algorithm is finished, otherwise, determine the search area of each control point, then go to step 2.

4 Experimental Results

To evaluate the proposed model and algorithm, two sets of experiments are presented. All the experiments are designed for extracting the boundary of ventricles from the brain MR images. The size of the image is 256*256, and the gray scales of the image are 256. In all experiments, the initialization of the model is



(a) (b) (c)

Fig. 3. Segmentation result of MRI brain image by our model. (a) initialization, (b) the result achieved at the fist stage using hierarchical deformation mechanism, (c) the final result by fine match algorithm.

provided by user. The first set of experiments demonstrates the performance of the segmental active contour model and its results at two processing stages. The second set of experiments compares our model with the classical snakes model[1] and GVF-Snakes[2].

The initialization of the model is shown in Fig.3(a). Figure 3(b) shows the result of the segmental active contour model at the pre-processing stage using hierarchical deformation mechanism, which is achieved generally within several iterations. The contour of model is close to the actual boundary and describes the approximate shape of brain ventricle. The accurate result is achieved using fine match scheme at the following stage, as illustrated in Fig.3(c). The model exactly conforms to the ventricle's boundary concavity.

Figure 4 gives the performance comparisons of the proposed model with the classical snakes and GVF-Snakes. All models use the same initial contour for segmenting the brain MR image, as shown in Fig.4(a). Using the classical snakes, the result is shown in Fig.4(b). In addition to being trapped by erroneous edges, this model has poor ability to move into the boundary concavity. Both GVF-Snakes and the proposed model have more accurate results, as shown in Fig.4(c) and Fig.4(d) respectively. However, the model presented in this paper utilizes the simpler and more efficient methods for computing internal energy, the convergence time is reduced. In case of segmenting images with a lot of fake edges around the actual boundary, our model may achieve better results, as it deforms depending on region-based method like region growing techniques instead of those edge-based methods.

5 Conclusion

In this paper, a segmental active contour model is presented for medical image segmentation. A key feature of our model is that it utilizes not only the edge-

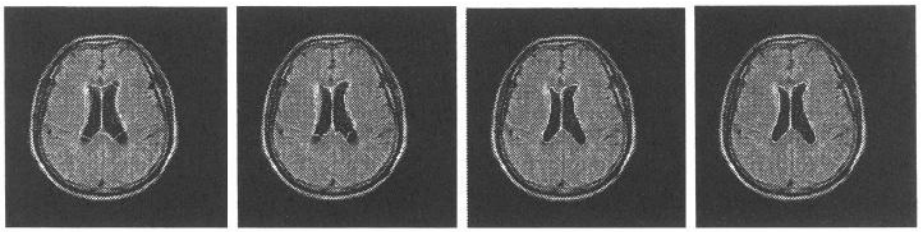


Fig. 4. Comparisons of our model with classical snakes and GVF-snakes. (a) initialization for all models, (b) the result of classical snakes, (c) the result of GVF-snakes, (d) the result of our model.

based method but also the region-based method. The edge-based method is used to find the approximate shape of object in a short time. The region-based fine match scheme is then adopted to localize the contour on object's boundary exactly. The experiments show that our new model is a competent approach for medical image segmentation.

References

1. Kass, M., Witkin, A., Terzopoulos,D.: Snakes: Active contour models. Int. Journal of Computer Vision. **1** (1987) 321–331
2. Chenyang Xu and Jerry L. Prince.: Snakes, Shape, and Gradient VectorFlow. IEEE Transactions on Image Processing. **7(3)** (1998) 359–369
3. D. Perrin and C. Smith.: Rethinking classical internal forces for active contour models. Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 2001.
4. Patrik Brigger, Jeff Hoeg and Michael Unser.: B-Splines: A Flexible Tool for Parametric Contour Detection. IEEE Transactions on Image Processing. **9(9)** (2000) 1484–1496
5. Dinggang Shen, C. Davatzikos.: An adaptive-focus deformable model using statistical and geometric information. IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI). **22(8)** (2000) 906–913
6. Z. Xue, S.Z. Li, and E.K. Teoh.: Ai-eigensnake: An affine-invariant deformable contour model for object matching. Image and Vision Computing. **20(2)** (2002) 77–84
7. Horace H S Ip and Dinggang Shen.: An affine-invariant active contour model (AI-snake) for model-based segmentation. Image and Vision Computing. **16(2)** (1998) 135–146
8. Remi Ronfard.: Region based strategies for active contour models. International Journal of Computer Vision. **13(2)** (1994) 229–251

A Level Set Algorithm for Contour Tracking in Medical Images

Yue Li and Qingyu Tang

Department of Biomedical Engineering, Tsinghua University, Beijing, China
liyue98@mails.tsinghua.edu.cn

Abstract. In this paper, a novel curve evolution strategy driven by boundary statistics for the segmentation of medical images is proposed and realized under the Level Set framework. It has a speed term similar to that of the Chan-Vese's method [1] for bimodal pictures, but is driven by boundary statistics (the statistics of intensity in an observing window) instead of the global statistics. In the case of multimodal pictures, the target's shape can, therefore, be more easily recovered. Here, we present methods for shape prediction based on the signed distance functions and extension field constructed from the boundary statistics. Employing the above techniques, our algorithm can adaptively handle both the sharp and smooth edges of the target, and its efficiency is demonstrated in the contour tracking of medical images.

Key words: curve evolution, level set, boundary statistics, contour tracking, shape prediction

1 Introduction

Algorithms based on curve evolution are classified as active contour. With the advantage of noise-insensitivity and its preservation of closed edge, it has brought much success to computer vision, especially in the challenging task of medical image segmentation. The first curve evolution model for image segmentation is the *snake,* proposed by Kass *et al* [2], a parametric contour. Sethian first proposed the geometric contour using level set. To deal with the problems of curve evolution, the curvature-dependent level set method, proposed by Osher and Sethian [3], has been used extensively. It allows for cusps and corners, is unchanged in higher dimensions and can naturally handle the topological changes in the evolving curve.

In the past decade, segmentation algorithms using geometric contour have focused on the evolution speed of the front or interface. Earlier level set techniques such as [4,5,6] only utilized the local strategy, they try to reduce the propagation speed where the image gradient is high. Compared with the local-based methods, regional statistics in curve evolution have prove advantageous, as they employ both the local and global information for pulling or pushing the front to the desired boundary. For example, the level set techniques based on Bayesian theory or clustering ([7,8]) are more robust than those mentioned previously. Most of them, however, need prior knowledge of the distribution of the

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 137–144, 2004.

image's intensity, and the estimations directly influence the propagating force and finally the outcome of the segmentation.

Chan *et al* [1] and Tsai *et al* [9] both proposed the curve evolution model based on the optimization of the Mumford-Shah function. By solving the region competition-based Mumford-Shah problem, one need not know the exact values of the regional statistics in advance and the initial curve could be arbitrary. The global bimodal method is simple but not suitable for most images, while the multimodal method is impractical. In addition, it becomes necessary to decide the number of modals in advance, which is a major drawback.

Another problem many level set methods face is the choice of the initial curve. Due to the CFL restriction, the front cannot move further than a single space step during each time interval. The position and the shape of the initial curve are therefore crucial to the computational period. If the initial curve is far from the target edge, a large number of iterations would be required in order to propagate the curve toward the target edge.

In this paper we will pay special attention to solving the following problems:

1. To avoid looking for a global minimum in the multimodal problem, we introduce boundary statistics to take the place of regional statistics for propagating the front.
2. In the contour tracking of 3D medical images, information of detected edges from previous slices can be used to predict edges in the current slice. The number of iterations needed to converge to the target edge can be reduced, and efficiency increases.

2 Method

2.1 Overview of Level Set Method

In the 2D case, the level set method assumes that on a planar region R, a closed curve $\boldsymbol{C}$ separates the region into two parts: the inside is R_1 and the outside is R_2 (Fig. 1). The curve is embedded in a higher dimensional function $\phi(x, y, t)$ and at time t, it is defined by the zero level set of ϕ. The *signed distance function* (SDF) of $\boldsymbol{C}$ is often chosen as its level set function. The absolute value of SDF at point A is equal to the distance between A and the nearest point on the curve to A, inside $\boldsymbol{C}$ it is taken as negative, and outside, positive. Imagine that $\boldsymbol{C}$ moves in directions normal to itself with a speed function F. The motion of $\boldsymbol{C}$ can be written as the following curve evolution equation:

$$\boldsymbol{C} = \boldsymbol{N}F \ . \tag{1}$$

where $\boldsymbol{N}$ is the unit normal vector of $\boldsymbol{C}$. Osher and Sethian [10] derived the partial differential equation for the level set function ϕ which matches its zero level set with the evolving front:

$$\phi_t = -F|\nabla\phi| \ . \tag{2}$$

given $\phi(x, y, t = 0) = \phi_0(t)$.

Equation (2) is the level set equation, an initial value formulation. Its numerical solution can be computed on the discrete 2D grids. The upwind scheme for finite difference approximations of the spatial and temporal derivatives is used and leads to a unique, entropy-satisfying solution with sub-grid resolution. Sethian [10] also describes a fast algorithm called *narrow band level set method.*

2.2 Boundary Statistics

Assume g is a gray scale picture defined on R and P is an arbitrary point on $\boldsymbol{C}$. Consider $N(P)$, a neighborhood of P. Naturally, it contains a part of R_1, and a part of R_2. Let R_{1P} be the intersection of $N(P)$ and R_1, R_{2P} be the intersection of $N(P)$ and R_2 as shown in Fig. 1. We can find the statistical characteristics of g in these two sub regions: let a_{1P} be the mean of g in R_{1P}, a_{2P} be the mean of g in R_{2P}, σ_{1P} be the variance of g in R_{1P} and σ_{2P} be the variance of g in R_{2P}. The functions a_{1P}, a_{2P}, σ_{1P} and σ_{2P} are defined on $\boldsymbol{C}$ and we call them boundary statistics.

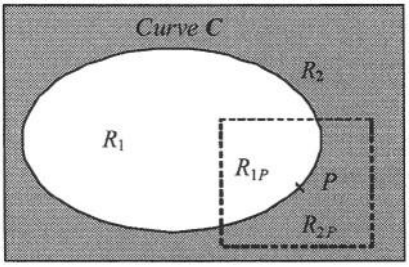


Fig. 1. The neighborhood *(dashed square)* of a point P on curve $\boldsymbol{C}$, including its two sub regions R_{1P} and R_{2P}

2.3 Shape Prediction

Given a target region R_1, or its boundary $\boldsymbol{C} = \partial R_1$, we can find the SDF, by performing a *signed distance transform* (SDT) on R_1, $d = \mathrm{SDT}(R_1)$, where d is defined on the image domain (x–y plane). For any R_1 (or closed curve $\boldsymbol{C} = \partial R_1$), there is a unique corresponding $d = \mathrm{SDT}(R_1)$ (or $d = \mathrm{SDT}(\boldsymbol{C})$). We note the region $R_1 = \{(x, y) | d < 0\}$ as $R_1 = \mathrm{SDT}^{-1}(d)$ (or note $\boldsymbol{C} = \{(x, y) | d = 0\}$ as $\boldsymbol{C} = \mathrm{SDT}^{-1}(d)$) and call it the 'inverse' transform of SDT.

Note $g_i(x, y), i \in N$, as an image sequence. A is the target object and R_{Ai} is the known target region of A in $g_i(x, y), i = 1, 2, \ldots, n-1$. The prediction problem is to find a region $\hat{R}_{An} = e(R_{A1}, R_{A2}, \ldots, R_{An-1})$ as an estimation of R_{An}, the unknown target region of A in $g_n(x, y)$. We propose to estimate the shape via SDF. We let $d_{Ai} = \mathrm{SDT}(R_{Ai}), i = 1, 2, \ldots, n-1$ be the SDF of R_{Ai}

and try to find a $\hat{d}_{An} = e'(d_{A1}, d_{A2}, \ldots, d_{An-1})$, as an estimation. We then let $\hat{R}_{An} = \mathrm{SDT}^{-1}(\hat{d}_{An})$ be the estimation of region R_{An}. The simplest form of the function e' is a linear prediction (estimation), that is to say:

$$\hat{d}_{An} = e'(d_{A1}, d_{A2}, \ldots, d_{An-1}) = w_1 d_{A1} + w_2 d_{A2} +, \ldots, + w_{An-1} d_{An-1} = w^T d \ . \tag{3}$$

where w is the weighting vector. Let the edge of predicted $\hat{R}_{An}$, which may be very close to the true target boundary, be the initial curve.

2.4 Extension Field of Boundary Statistics

Given the well-segmented R_{An-1}, we can find the boundary statistics of the previous slice $g_{n-1}(x, y)$: a_1, a_2, σ_1 and the SDF d_{An-1}. Based on these, we construct the extension field $F(a_1)$, $F(a_2)$ and $F(\sigma_1)$ (note: not the propagation speed *F)* from the boundary statistics. The relation between an extension field and the corresponding boundary statistic x satisfies the following:

$$\begin{cases} F(x) = x, & d_{An-1} = 0 \\ \nabla F(x) \cdot \nabla d_{An-1} = 0, & d_{An-1} \neq 0 \end{cases} \ . \tag{4}$$

where x can be a_1, a_2 or σ_1. Readers can refer to the construction of extension velocity in Ref. [10].

2.5 Our Contour Tracking Algorithm

Our level set equation for contour tracking is:

$$\phi_t = -\alpha((g - F(a_2))^2 - (g - F(a_1)))^2 |\nabla\phi| + \gamma\kappa|\nabla\phi| - \beta \cdot s(a_1, a_2, \sigma_1) \cdot (\phi - \phi^*) \ . \tag{5}$$

where $s(a_1, a_2, \sigma_1) = \exp\left[-\lambda\left(\frac{F(a_2) - F(a_1)}{F(\sigma_1)}\right)^2\right]$. The first term on the right side is similar to the Chan-Vese's method, but is driven by the extension field of the boundary statistics rather than by global statistics. The second term is the curvature term, where κ is the curvature of the level set function. We set the SDF d_{An-1} as ϕ^* to be a shape constraint. The term $s \in (0, 1]$ is a controlling factor: since $|(F(a_1) - F(a_1))/F(\sigma_1)|$ becomes greater where the boundary character is strong, weakening the constraint, and becomes small where the boundary character is weak, enforcing the constraint, it can adaptively handle both the sharp and blurred edges. α, β, γ and λ are the coefficients, controlling the weight of every term.

In the contour tracking algorithm, we first calculate the predicted shape as the initial curve, then construct the extension fields, and finally evolve the curve using the fields and local intensity. Fig. 2(d)-(f) illustrate the predicting phase, Fig. 2(g) and (h) show the extension fields, $F(a_1)$ and $F(a_2)$, constructed from the boundary statistics of g_2 and R_{A2}, and Fig. 2(i) is the tracking result. Considering the inhomogeneity of intensity across different slices, a correction factor Δ should be added to $F(a_1)$ and $F(a_2)$, which is equal to the difference between the means of g_2 and g_3 in region $\hat{R}_{A3}$.

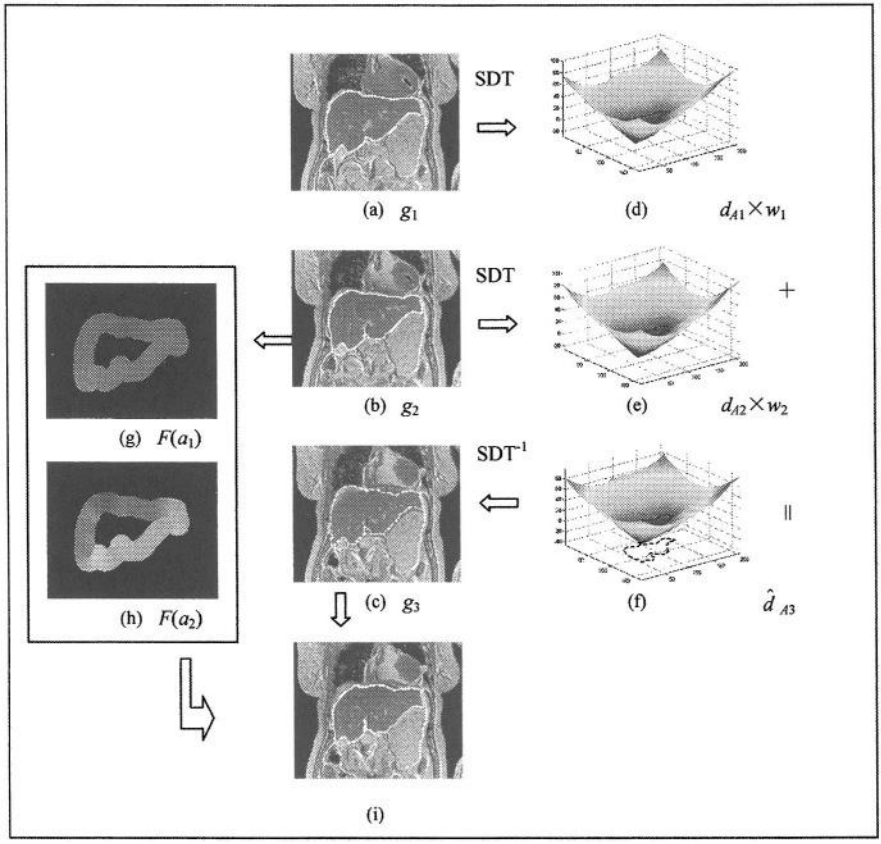


Fig. 2. Flow chart of our contour tracking algorithm

3 Result

We applied the above strategy to the MRI liver data from the Visible Human Dataset. The size of each slice is 207 × 167 pixels, with a total of 39 slices, and 8-bit gray scale. The neighborhood of any point is chosen as a 16 × 16 square window centered around the point. The coefficients in the level set equation are: $\alpha = 1$, $\beta = 0.28$, $\gamma = 0.02$ and $\lambda = 1$. We chose a simple form of the linear prediction coefficients: $w_{n-2} = -1$ and $w_{n-1} = 2$. Fig. 3 shows the results of the 29th slice. We can see that the predicted edge is very close to the true edge. In most slices, it takes no more than 10 iterations to converge to the target boundary. Inhomogeneity across slices can be handled with the correction factor Δ. Manual assistance was needed only in one slice (18th) in order to achieve the correct result. Fig. 4 shows the 3D distribution of the contours, and gives the reconstructed surface.

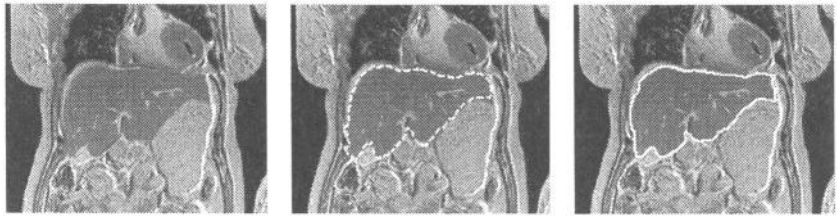


Fig. 3. The 29th slice from 3-D liver images: left is the origin image; middle is the predicted initial curve *(dashed line);* right is the extracted contour *(white line)*

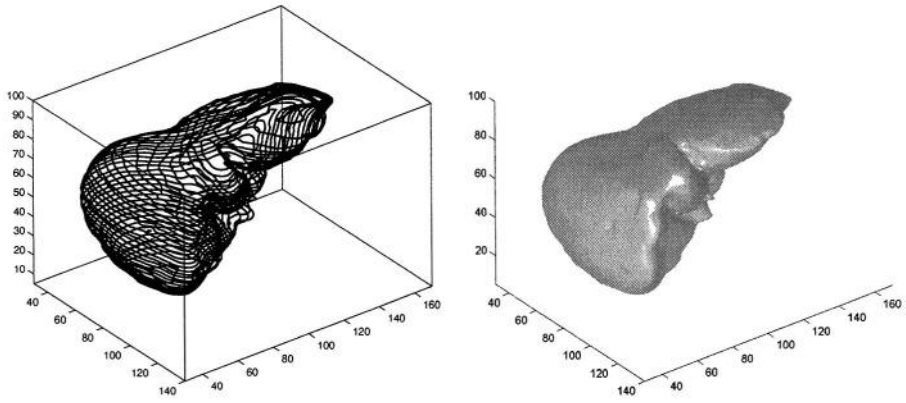


Fig. 4. Left: the 3-D contour set of liver. Right: the reconstructed surface from the contour set

In Fig. 5, results are shown for comparison. The images in the left column are the original images of the 26th and the 37th slice; the middle column shows the results obtained by using the gradient-based method proposed in Ref. [6]; the right column shows the results obtained by using our algorithm. From this comparison, it is evident that our method can lead to more accurate results.

To measure the similarity between the predicted shape and the final result, we define two indicators : s_1 and s_2. For R_{An} and $\hat{R}_{An}$, $s_1 = |(A_{\text{common}})/A|$, and $s_2 = |(A_{\text{diff}})/A|$, where A_{common} is the area of $\hat{R}_{An} \cap R_{An}$ (intersection), A is the area of R_{An} and A_{diff} is the area of $\hat{R}_{An} \oplus R_{An}$ (exclusive or). The mean of s_1 measured in this experiment is 0.96, and the mean of s_2 measured is 0.11, while the means of the similar indices for R_{An-1} and R_{An} are 0.93 and 0.13, showing the validity of our prediction method.

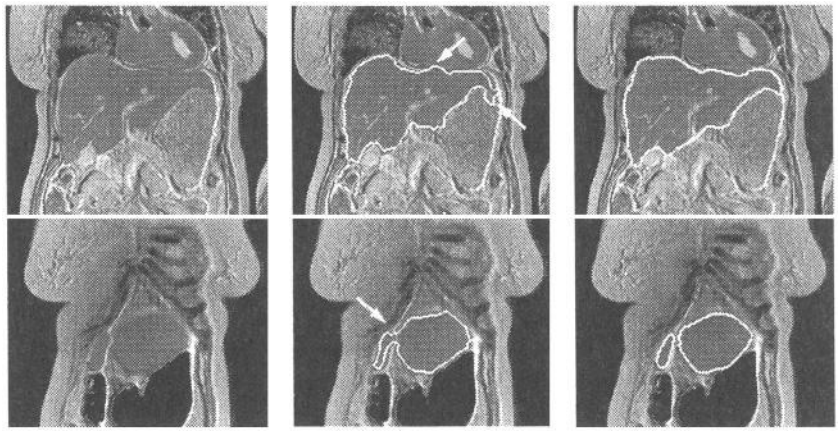


Fig. 5. Left column: original images. Middle column: results obtained by using the gradient-based method in [6]. Right column: results obtained by using our algorithm. Arrows are used to point out the differences between the results of the two methods

4 Discussions and Conclusion

Our contribution in this paper is the use of boundary statistics in curve evolution and the application of SDF to shape prediction. The level set equation in this paper is partly derived from the Chan-Vese's method. The original Chan-Vese's method is designed to handle bimodal pictures. Our method can handle not only bimodal but also multimodal pictures, depending on the size of the observing window. In an extreme case, if the observing window is extended to the whole image domain, the boundary means a_1 and a_2 will be equal to the corresponding global statistics in their method. In our method, the boundary statistics are shown to be obtained from an observing window and vary along the curve, so it can also be viewed as a non-linear filtering method or an adaptive thresholding process. Since SDF has been an effective way to describe shapes, even a simple use of it can work well. The set of SDFs is not a linear space, since in most situations, combinations of SDFs do not produce a SDF. However, the statistical process of SDF and algebraic operation on SDF are still available. The validity of such a use can be seen in the works of Leventon *et al* [11] and Tsai *et al* [12].

Acknowledgements

We would like to sincerely thank Prof. Shangkai Gao for her instruction and Mr. Gusphyl Justin for revising this paper. We also thank the National Library of Medicine and the Visible Human Project for providing the Visible Human Dataset.

References

1. T.F. Chan., L.A.Vese, Active contours without edges, IEEE Trans. Image Processing, 10(2), Feb. 2001 pp: 266-277
2. M. Kass, A. Witkin, and D. Terzopoulos, Snakes: Active contour models, Int. J.Comput. Vis., 1, 1987, pp: 321-331.
3. S. Osher and J. A. Sethian, Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi Formulation, J. Comput. Phys., 79(1), 1988, pp: 12-49.
4. V. Caselles, F. Catte, T. Coll, and F. Dibos, A geometric model for active contours, Numer. Math., 66(1), 1993, pp: 1-31.
5. R.Malladi, J. A. Sethian, and B. C. Vemuri, Shape modeling with front propagation, IEEE Trans. Pattern Anal. Machine Intell., vol. 17, Feb. 1995, pp: 158-175.
6. R. Malladi and J. A. Sethian, An O(N log N) algorithm for shape modeling, Appl. Math., Proc. Nat. Academy Sci., 93(18), Sept. 1996, pp: 9389-9392.
7. J. S. Suri, Leaking prevention in fast level sets using fuzzy models: An application in MR brain, in Proc. Int. Conf. Inform. Technol. Biomedicine, Nov. 2000, pp: 220-226.
8. X. Zeng, L. H. Staib, R. T. Schultz, and J. S. Duncan, Segmentation and measurement of the cortex from 3-D MR images, Med. Image Comput. Computer-Assisted Intervention, 1998, pp: 519-530.
9. A Tsai., A.Yezzi Jr., A.S. Willsky, A curve evolution approach to smoothing and segmentation using the Mumford-Shah functional, IEEE Conf. Computer Vision and Pattern Recognition 2000, vol.1, 13-15 Jun. 2000, pp: 119-124.
10. J.A. Sethian, Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science, 2nd ed., Cambridge University Press, 1999.
11. M.E Leventon, W.E.L. Grimson, O.Faugeras, Statistical shape influence in geodesic active contours, IEEE Conf. Computer Vision and Pattern Recognition 2000, vol. 1, 13-15 June 2000, pp: 316-323.
12. A. Tsai, A. Yezzi Jr., W. Wells *et al,* A shape-based approach to the segmentation of medical imagery using level sets, IEEE Trans. Medical Imaging, 22(2), Feb. 2003, pp: 137-154.

Robust Object Segmentation with Constrained Curve Embedding Potential Field

Gary H.P. Ho and Pengcheng Shi

Biomedical Research Laboratory
Department of Electrical and Electronic Engineering
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
{garyho,eeship}@ust.hk

Abstract. We have earlier introduced an implicit vector field representation for arbitrary number of curves in space, the curve embedding potential field (CEPF), and a general image segmentation strategy based on the detection of the CEPF distortion under the influence of vector-form image data [3]. In this paper, we present an improved CEPF framework which incorporates prior knowledge of the object boundary and has consistent object definition through a region growing process. The embedded implicit curves deform through the image- and model-induced changes of the CEPF, which evidently improves the segmentation accuracy under noisy and broken-edge situations. Further, the closure enforcement and the natural advection on the curves enhance the stability of CEPF evolution and the implementation is straightforward. Robust experimental results on cardiac and brain images are presented.

1 Introduction

A generic description of an object would be a closed contour formed around the object boundary edges. Since the original active contour model [5], many researchers have contributed to the improvement of the *Snakes*-type boundary segmentation paradigm by achieving better balance between the necessary structural constraints to maintain model integrity and the ability to fully utilize the image information [2, 7]. More recently, the geodesic active contour strategy, which combines the active contour models and the level set methods, uses geometric representation of the boundary model and thus permits greater curve deformation and topological changes [1, 6]. This improved model evolves like a waterfront that propagates along the normal direction until the whole front is blocked by object edges or reaches the border of the image space.

We have earlier proposed an implicit curve representation and segmentation strategy [3], with inspirations from the vector form level set formulation [8]. The key contribution has been that continuous curves are implicitly represented by a potential field in space, the curve embedding potential field (CEPF), and all operations are in the native vector form of edge data and CEPF (see Fig. 4 for the interaction between the gradient vector flow and CEPF). Being geometric

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 145–153, 2004.

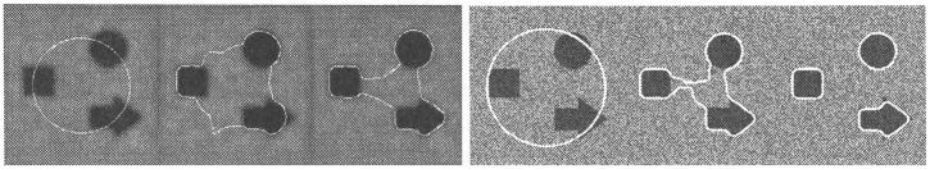


Fig. 1. Problems with the original CEPF (left): curve stabilized prematurely when GVF between separated objects are weak. Improved segmentation results on speckle noise-corrupted image (12.5% variance), with closure enforcement and natural curve speed but no prior edge model (right).

in nature, CEPF allows merge and break of the underlying contours and offers better handling of discontinuities. Object segmentation is achieved by iterative vector field construction, vector data-driven field evolution and regularization, and the detection of the CEPF vector crossings [3] . We have also shown that the CEPF strategy exhibits improvements on segmentation accuracy over the level set strategies on certain medical images.

1.1 Several Remaining Issues of the Original CEPF

GVF as Data Source: Since the gradient vector field (GVF) has a continuous domain of influence, with the edge points appearing as vector sinks [9], GVF has been used in the CEPF segmentation strategy as the guiding data source [3]. One problem of the original CEPF framework relates to the GVF magnitude distribution for complicated image scene. Although not always in strict sense, GVF vectors belonging to the different sides of the same edge are typically opposite to each other, and the CEPF will slide on the GVF field to check for edge existence. However, because locations which are far away from high gradient areas receive very little diffusion in the GVF computing process, the GVF magnitudes at these points are indeed very small. Hence, in these locations the CEPF field will have very little GVF-induced changes in the evolution process, at least numerically, and the evolution may stop prematurely, before reaching the true boundaries (see Fig. 1 left). In the cases where images are not very noisy, normalized GVF field has been used as the data guidance to speed up the evolution [3]. On the other hand, amplified noise power from the GVF normalization process may cause further problems.

Moreover, even if the noise level is low, when detecting multiple objects, contour between objects will degenerate into a stabilized line (like a tail between regions). In level set method, this line will automatically disappear because the ϕ value of both sides of the contour have the same sign (either both > 0 or both < 0) and will not be extracted out. In CEPF method, this line will remain even if both left and right sides belongs to the image background.

Besides, CEPF evolution does not require the embedded contour to be closed curve. However, for content rich or noisy images, the resulting GVF contains

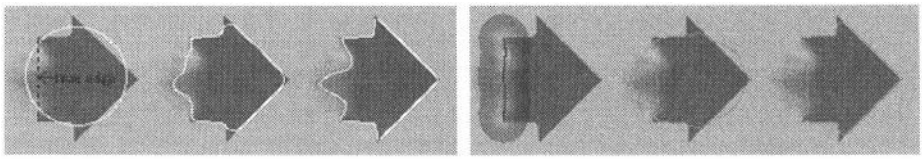


Fig. 2. Remaining problems even with curve closure: mis-segmentation because of occluded edge. The dotted line in the leftmost image is the true boundary and the left three images show the segmentation based on the blurred edge. The vertical line in the forth image is an additional user input and the overlayed dark blob is an auxiliary probability model created by diffusing the user input line. The rightmost image is the segmentation using model-constrained CEPF on synthetic image.

many undesired singular locations so that it is hard to guarantee the continuity of the local maximum and results in generally non-closed curves. Since the original CEPF formation enables natural interpolation among the embedding curve elements, these open-ended curve pieces extends everywhere as the CEPF is iteratively reconstructed. The resulting CEPF may form a web that divides the image into many regions, instead of given out specific object segments.

Occlusion and Weak Edges: Additional problems are caused by weak or occluded edges in an image, as shown in Fig. 2. To get the improved results of Fig. 1, the modified term (in Equation 4 and 5) for low GVF diffusion imposes a natural advection force on the evolving front, which helps to break down the curve into several closed curves for detecting multiple objects. However, the same effect also applies to the missing or occluded edge along object boundary, which means that the moving front will pass through those areas wherever no GVF force is available.

1.2 Solution Overview

To deal with issues related to the GVF data source and curve closure, we have incorporated a modified term in the CEPF updating equation (in Equation 4 and 5) and a region growing procedure during curve detection (see flowchart in Fig. 3) for speeding up, tails elimination, and closure enforcement in the CEPF evolution process. Details are discussed in the later sections and one improved result is shown in Fig. 1 right. In order to distinguish between empty space and missing/occluded edge, we opt to incorporate an edge probability distribution as a prior knowledge during segmentation so that, instead of the unconditional natural advection, the prior will take effect if no image data (GVF) presents.

2 Constrained CEPF Evolution for Object Segmentation

The constrained CEPF algorithm is an iterative process which starts from one or more initial contour(s) and evolves until converges to closed contours along

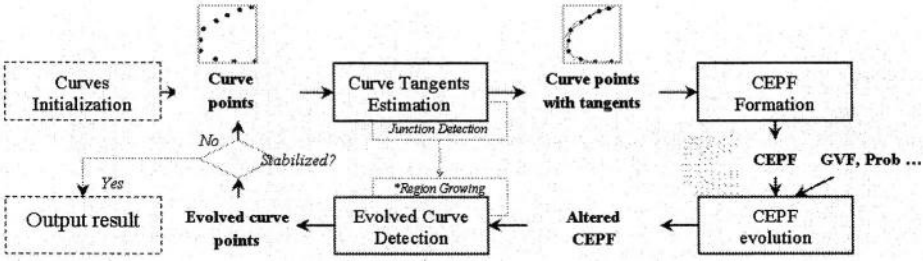


Fig. 3. Flowchart of the model-constrained CEPF segmentation method. Boldface letters are the input and output between major steps represented by rectangular blocks.

object boundaries. The central idea is that the contour model is implicitly represented by a potential field where locations with zero potential represent the embedded curve. Instead of moving the curve elements directly, we first acquire the CEPF representation through a curve-to-field transformation, and then trigger a CEPF distortion by applying prior- and image-derived, vector-form edge information on the CEPF. As a result, physical curve elements implicitly propagate towards object boundaries, while structural constraints will be loosely defined in the formation of the CEPF so that the length, shape, and even the number of embedded curves can be changed during CEPF evolution. All procedures are shown in Fig. (3), we here only describe those modified blocks that different from the original CEPF framework. See [3] for remaining blocks.

Curve Initialization: Initial curves can be placed by default, by user interaction, or through the prior knowledge of the boundary distribution. For example, if the edge probability map (a distribution with higher value for frequent boundary occurrence at particular location) is available, we can transform the map into vector field using GVF method and then apply native CEPF on the vector field to obtain an initial curve position. The edge maps in Fig. 5 are created by applying gaussian diffusion on hand-traced contour.

Junction Detection: As stated earlier, too many curve pieces are undesired side effects from noisy images. Hence, we control the topology of the curve pieces by applying region growing in the *evolved curve detection* step to ensure that the curve points maintain closed loops. Region growing is computationally expensive compared to other procedures in the framework, which are limited operations within a narrowband, and thus is not preferred to be included in every iteration. Due to the interaction between CEPF, GVF, and prior edge probability, unwanted curve segments usually intersect *desired* curves and form *T* or *Y* junctions. Therefore, existence of these types of junctions can be indicators of having unwanted curves which then triggers the region growing routine. Hence, we include an additional junction detection metric as following. Let $\mathbf{z}_3$ be the third curve neighbor of **z,** $\mathbf{t}$ = *Tangent*(**z**), and **n** be the unit normal of **z.** For $1 \leq i, j \leq 3$ and $i \neq j$, define

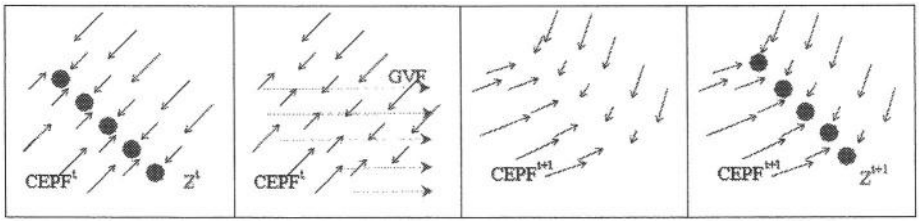


Fig. 4. A simple illustration of the effect of GVF on CEPF. The two leftmost figures display the initial configuration before evolution, where the red points are the set of curve elements Z, the blue arrows are the CEPF vectors (scaled for better illustration), and the long green dotted arrows represent the GVF field. The third figure shows the CEPF after the influence of GVF and the fourth is the result of curve detection on the evolved CEPF.

$$G_{\mathbf{z}} = \{\mathbf{z}_i, \mathbf{z}_j\} \quad \text{when } ((\mathbf{z}_i - \mathbf{z}) \cdot \mathbf{t})((\mathbf{z}_j - \mathbf{z}) \cdot \mathbf{t}) > 0 \tag{1}$$

$$Jnt(\mathbf{z}) = \begin{cases} 1, \text{ if } (\mathbf{z}_3 - \mathbf{z}) \cdot \mathbf{t} = 0 \text{ or} \\ \quad ((\mathbf{z}_i - \mathbf{z}) \cdot \mathbf{n})((\mathbf{z}_j - \mathbf{z}) \cdot \mathbf{n}) < 0, \\ \quad \{\mathbf{z}_i, \mathbf{z}_j\} \in G_{\mathbf{z}} \\ 0, \text{ otherwise} \end{cases} \tag{2}$$

where set $G_{\mathbf{z}}$ contains two neighbors of $\mathbf{z}$ lying on the same side of the normal vector $\mathbf{n}$ and $Jnt(\mathbf{z})$ denotes the junction flag. Point $\mathbf{z}$ is likely to be a junction if $\mathbf{z}_i$ and $\mathbf{z}_j$ lie on the opposite sides of the tangent $\mathbf{t}$, implying that $\mathbf{z}_1$, $\mathbf{z}_2$, and $\mathbf{z}_3$ locate in three completely different quadrants relative to the local intrinsic coordinates of $\mathbf{z}$ (formed through $\mathbf{t}$ and $\mathbf{n}$). This situation is unlikely to occur along smooth curve.

In our present method, junction detection and region growing processes are separated. Junction detection is performed when doing curve tangents estimation because neighbor searching is also required. And region growing is deferred to the curve detection step, because sometimes segmented regions are preferred as the outputs so that we can provide both boundaries and regions.

Model-Constrained CEPF Evolution: Suppose that we have an *imaginary* curve point at location $\mathbf{x}$ and it moves with an image-derived GVF velocity $\mathbf{g}(\mathbf{x})$, then new position after Δt time is approximately $\mathbf{x} + \Delta t\, \mathbf{g}(\mathbf{x})$. Of course this change of curve position is completely image-based. Now, we judge the reliability of this movement based on prior knowledge of the curve location. Assume that the prior edge probability distribution is $prob(\mathbf{x})$, the difference of edge likelihood between the original and new positions is then $prob(\mathbf{x}) - prob(\mathbf{x} + \Delta t\, \mathbf{g}(\mathbf{x}))$. Negative value of this probability difference means that the image and the prior information agree that the curve evolution direction is heading for the possible boundary. Otherwise, if the two information contradicts to each other, we want to slow down the movement according to the probability difference. To do this, we model the difference as a resistance r to the curve motion, which is used to construct a scaling factor of the velocity $\mathbf{g}(\mathbf{x})$:

$$r(\mathbf{x}) = max\Big(prob(\mathbf{x}) - prob\big(\mathbf{x} + \Delta t\, \mathbf{g}(\mathbf{x})\big), 0\Big) \tag{3}$$

$$\mathbf{u}(\mathbf{x}) = \frac{e^{-r(\mathbf{x})} - e^{-1}}{1 - e^{-1}} \mathbf{g}(\mathbf{x}) + \epsilon\, \mathbf{n}(\mathbf{x}) \tag{4}$$

where $\mathbf{u}$ is the image-induced and model-constrained speed for an *imaginary* curve point at $\mathbf{x}$. In Equation 3, r ranges from 0 to 1 with 0 meaning no resistance for a curve point leaving location $\mathbf{x}$. In Equation 4, r is mapped to a exponential decay function which is then used as a scaling factor for the driving force $\mathbf{g}$, vector $\mathbf{n}$ is the normal direction of the curve at $\mathbf{x}$, and the constant $\epsilon \ll \Delta t$. These terms are used to compensate the problem of small GVF magnitude if the native GVF is used as the data source. Note that $\mathbf{x}$ is not required to be on the grid so that interpolations are needed for getting $\mathbf{g}(\mathbf{x})$ and $prob(\mathbf{x})$.

We describe $\mathbf{x}$ to be *imaginary* because we do not use finite number of points to represent the curve, i.e. a set of $\mathbf{x}$ does not exist. But we have the CEPF field that pointing towards the curve. Therefore, we can approximate $\mathbf{x}$ by $\mathbf{p} + CEPF(\mathbf{p})$ and the CEPF evolution with induced speed $\mathbf{u}$ is defined through:

$$\mathbf{d}^{t+1}(\mathbf{p}) = CEPF^t(\mathbf{p}) + \Delta t\ \mathbf{u}(\mathbf{p} + CEPF^t(\mathbf{p})) \tag{5}$$

$$\mathbf{t}^{t+1}(\mathbf{p}) = \Big(\mathbf{p} + \mathbf{d}^{t+1}(\mathbf{p})\Big) - \Big(\mathbf{p}_i + \mathbf{d}^{t+1}(\mathbf{p}_i)\Big) \tag{6}$$

$$CEPF^{t+1}(\mathbf{p}) = \mathbf{d}^{t+1} - \Big(\mathbf{d}^{t+1}(\mathbf{p}) \cdot \mathbf{t}^{t+1}(\mathbf{p})\Big) \frac{\mathbf{t}^{t+1}(\mathbf{p})}{|\mathbf{t}^{t+1}(\mathbf{p})|^2} \tag{7}$$

In Equation 5, $\mathbf{d}^{t+1}$ is the new curve location relative to $\mathbf{p}$, and Δt is the time step that governs the speed of the evolving CEPF. In Equation 6, $\mathbf{t}^{t+1}$ requires a proper reference neighbor $\mathbf{p}_i$ of $\mathbf{p}$, but it would be very inefficient to determine $\mathbf{p}_i$ for each $\mathbf{p}$ [3]. Hence, we can either take the approximation $\mathbf{f}^{t+1} \approx \mathbf{d}^{t+1}$ or $\bar{\mathbf{t}}^{t+1}(\mathbf{p}) \approx -Laplacian\big(\mathbf{g}(\mathbf{p} + CEPF^t(\mathbf{p}))\big)$. Equation 6 and 7 are used to keep CEPF vectors normal to the embedded curve so that, if Δt is small enough, this orthogonality property and the CEPF re-formation in the next iteration ensure the embedded curve does not have abrupt changes (see Fig. 4). Therefore, the structure of the curves is implicitly maintained.

Region Growing for Closure Enforcement: Discrete 2D region growing is trivial but rounding real valued curve points may create gaps along the discretized curves, failing the region growing process. Existence of tails and multiple curves makes the implementation of contour following unpleasant. Observing that the narrow band (NB) CEPF around the curve is actually a dilated version of the evolving front and the band itself has already been discretized, thus applying region growing on CEPF narrow band would be a natural solution. Precise tail-removal morphological operations are defined as following:

$$Mask = \{(i,j) \mid -w \le i,j \le w\, \forall\, i,j \in \mathbb{N}\} \tag{8}$$

$$NB = Z \oplus Mask \tag{9}$$

$$Z' = Z \cap ((Region(NB)) \oplus Mask) \tag{10}$$

Here, Z is the set of curve points extracted, $\oplus$ denotes the dilation operation, integer w represents the narrow band width, $Mask$ is a set of point within a $w \times w$

square window, Z' is the resulting tail-free set of curve points, and *Region*() is the region growing operation which outputs the set of points bounded by the narrow band.

3 Experiments and Discussions

Improved segmentation results from medical images are shown in Fig. 5. In Fig. 5, we can observe from the difference between the results (last column) and the prior edge probability distribution (user-input column) that image boundary are basically tracked if they are clear. This shows that image data is taking the major effect. The 1^{st} row in Fig. 5 is boundary-based segmentation of white matter in a brain image. The boundary between white and grey matter are quite clear but other kind of edges, such as the skull, make the contour hard to initialize. The mean prior shape (blue lines) in this example is created by manual thresholding the pixel intensities and this prior greatly helps the initialization. The segmentation result (contour in black) together with the prior shape (contour in blue) are shown in the rightmost figure of the 1^{st} row. The 2^{nd} and 3^{rd} rows give an example of changing the initializations for the same image. Without the prior, better initializations are not enough to produce good segmentation. Effect of occluded or irrelevant edges are regularized by using prior information. The 4^{th} row shows a situation of small boundary separation between neighboring objects where edges are smoothed out during the process of GVF [4]. Although the prior does not aligned with the object edges, the existence of two prior contours forces the CEPF to be separated.

For simplicity, we have used the same prior probability density as both the curve initialization and the prior data force for curve evolution. This can be easily done by first using normalized GVF of the probability density as the only data input of our CEPF algorithm until the curve stabilized to become a model-based initial curve. The CEPF algorithm is also applicable to probability density because we only need a vector field that shows the most likely direction to the nearest edge. The underlying meaning of the gradient image and edge probability are very similar since they both provide a boundary likelihood in space. The difference would be the source of information. So, the result of taking GVF from either one would be similar but the probability density is preferable for initial boundary estimation since it usually gives a unique closed band around the target so that the evolution of CEPF on texture regions might be avoided. After the initial curve is obtained, curve evolution moves on by changing the data input as the integrated field defined in Equations 3 and 4. The image sequences shown in Fig. 5 can be viewed as a coarse to detail curve refinement using CEPF method with increased complexity of input data.

Acknowledgement: This work is supported in part by Hong Kong Research Grant Council under CERG-HKUST6031/01E.

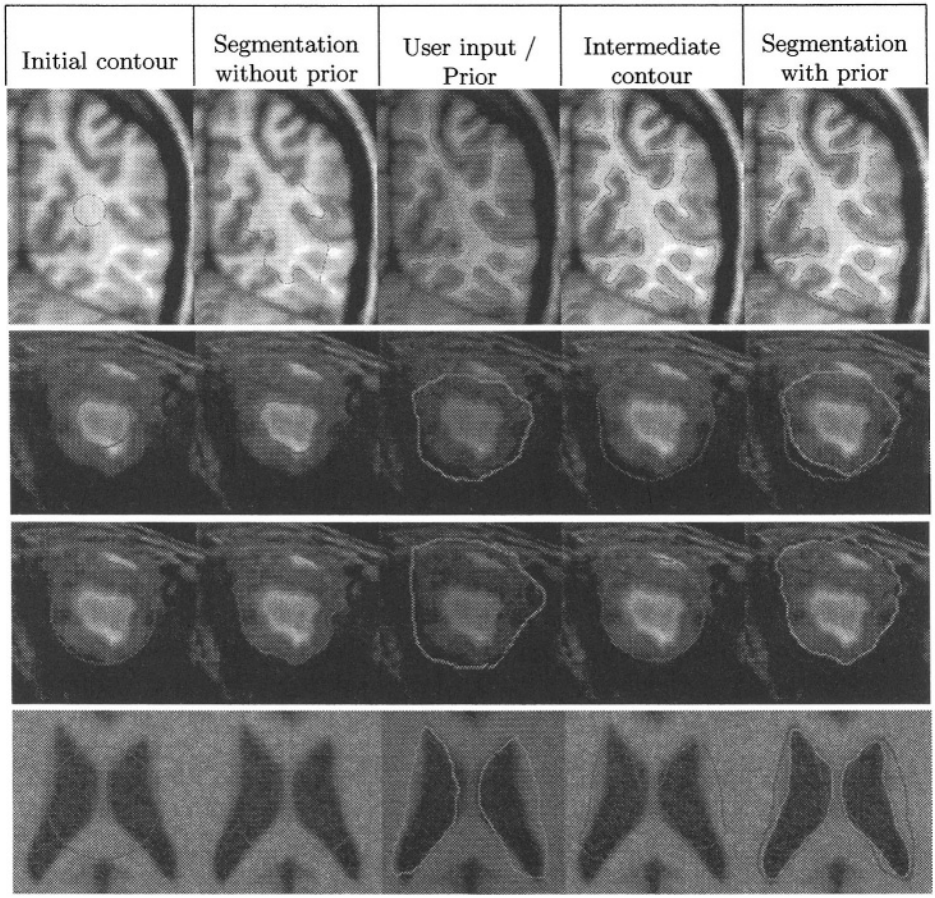


Fig. 5. Examples of model-constrained CEPF segmentation. In each case, initial boundary is first estimated using prior probability and the evolution continues with the integrated prior and image data as the driving force.

References

[1] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision,* 22:61–79, 1997.

[2] L.D. Cohen. On active contour models and balloons. *CVGIP: Image Understanding,* 53:211–218, 1991.

[3] H.P. Ho and P. Shi. Boundary finding with curve embedding potential field. In *6th Medical Image Computing and Computer Assisted Intervention,* 2003.

[4] H.P. Ho and P. Shi. Domain partitioning level set surface for topology constrained multi-object segmentation. In *IEEE International Symposium on Biomedical Imaging (ISBI),* 2004.

[5] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision,* 1:321–331, 1987.

[6] R. Malladi, J. A. Sethian, and B. C. Vemuri. Shape modeling with front porpagation: a level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 17(2):158–175, February 1995.

[7] T. McInerney and D. Terzopoulos. Topologically adaptable snakes. In *Fifth IEEE International Conference on Computer Vision,* pages 840–845, 1995.

[8] G. Ventura, J.X. Xu, and T. Belytschko. A vector level set method and new discontinuity approximations for crack growth by EFG. *International Journal for numerical methods in engineering,* 54:923–944, 2002.

[9] C. Xu and L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing,* 7(3):359–369, 1998.

Tracking Lumbar Vertebrae in Digital Videofluoroscopic Video Automatically

Shu-Fai Wong[1], Kwan-Yee Kenneth Wong[1], W.-N. Kris Wong[2], C.-Y. John Leong[2], and D.-K. Keith Luk[2]

[1] Department of Computer Science and Information Systems, The University of Hong Kong, Hong Kong
{sfwong, kykwong}@csis.hku.hk
[2] Department of Orthopaedic Surgery, The University of Hong Kong, Hong Kong

Abstract. Low back pain becomes one of the significant problem in the industrialized world. Efficient and effective spinal motion analysis is required to understand low back pain and to aid the diagnosis. Videofluoroscopy provides a cost effective way for such analysis. However, common approaches are tedious and time consuming due to the low quality of the images. Physicians have to extract the vertebrae manually in most cases and thus continuous motion analysis is hardly achieved. In this paper, we propose a system which can perform automatic vertebrae segmentation and tracking. Operators need to define exact location of landmarks in the first frame only. The proposed system will continuously learn the texture pattern along the edge and the dynamics of the vertebrae in the remaining frames. The system can estimate the location of the vertebrae based on the learnt texture and dynamics throughout the sequence. Experimental results show that the proposed system can segment vertebrae from Videofluoroscopic images automatically and accurately.

Key words: Motion Tracking, Spinal Motion Analysis

1 Introduction

Low back pain is one of the most common health disorders and its cost is enormous [1]. There is a general consensus that the diagnosis and the treatment of low back pain can be aided by analysing spinal movement [2]. Thus, spinal measurement techniques have been studied widely. At present, Videofluoroscopic imaging provides an effective method of obtaining images for spinal motion analysis. Generally, landmarks of a moving vertebra will be extracted from Videofluoroscopic video and will then be analysed. Landmarks are usually the corners of the moving vertebra and are usually extracted manually [3]. Unfortunately, the analysis is difficult and time consuming due to the low quality of the Videofluoroscopic images. Figure 1(a) shows typical Videofluoroscopic image of spine. Thus, a wide range of researches on automatic extraction of landmarks have been conducted, such as [4].

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 154–162, 2004.

In general, there are two main approaches in videofluoroscopic analysis. The first one is based on template matching and correlation (e.g. [5]) which is simple to implement and easy to understand. However, such approach involves pixel-to-pixel comparison and thus susceptible to changing contrast and pixel intensity of the image. Another approach which is based on feature detection is adopted in current research. Features can be corners, edges and shape. In [6], the vertebrae in the images are located by matching corners. In [7], active shape models is used to improve the robustness by introducing shape constraints. To reduce the searching size, generalized hough transform is used in [8]. Such approach is computational efficient but making unrealistic assumption of high image contrast. Edges and features have to be manually enhanced and refined before feature location can be done. It seems that most of the commonly adopted approaches can be considered as computer-aided but not automatic.

In this paper, we propose a method in which an active contour (or snake) will attach to the vertebrae automatically throughout the video sequence. Users only need to define landmark positions on the first videofluoroscopic image. The active contour formed from such landmarks will attach to the vertebra automatically in the remaining video images. This greatly reduces the effort of physicians in setting accurate landmarks of vertebra manually in every video frame. The reduction in human intervention means the reduction in error rate due to fatigue of the operator. Analysis on spinal motion can be done much more effectively and accurately.

2 System Architecture

The whole system consists of three major modules, namely feature learning module, feature detection module and tracking module. The workflow of the system is shown in figure 1(b).

Given the first image and the exact position of the landmarks, the feature learning module will learn the texture pattern encoded by Markov Random Field (MRF) [9] using Support Vector Machine (SVM) [10] along the edge. On the other hand, a snake or active contour [11] is formed from the landmarks. When second image is input, the feature detection module will detect the edge using the texture information from feature learning module along the snake. The snake will then fitted toward features (or edges) detected. The tracking module will then learn the dynamic of those landmarks using Kalman filter [12]. It will predict the location of the landmarks in the next frame and shift the snake accordingly. At the same time, the feature learning module will learn the texture pattern again. The feature detection will then detect the features in the next frame as described above and the whole process repeats. Thus, given the videofluoroscopic video, the corresponding time series data on the location of the landmarks will be obtained by the system.

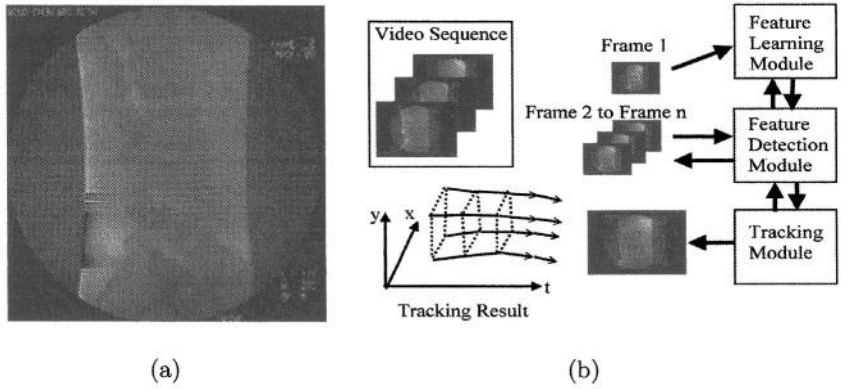


Fig. 1. *In (a), it shows a typical videofluoroscopic image of spine. In (b), it shows the workflow of the system.*

3 Implementation Details

As described in previous section, there are several components within each module in order to achieve the aim of automatic vertebrae tracking. These components are shared among the modules and work collaborately. These components include the MRF texture descriptor, the SVM pattern recognizer, the snake and the kalman filter. The implementation details of these components will be explored in this section.

3.1 Texture Description by Markov Random Field

Markov Random Field was first developed for texture analysis, e.g. [13]. It can be used to describe a texture and make prediction on the intensity value of a certain pixel given the intensity value of its neighborhood. The theories related to Markov Random Field can be found in [9].

In Markov Random Field, the neighborhood is defined as clique elements. Consider that $S = \{s_1, s_2, .., s_P\}$ is a set of pixels inside the image, and $N = \{N_s | s \in S\}$ is the neighborhoods of the set of pixels. In the system, the neighborhoods are the 8 pixels that with chessboard distance 1 away from the target pixel.

Assuming $X = \{x_s | s \in S\}$ is the random variables (the intensity value) for every pixel inside an image, where $x_s \in L$ and L = {0,1, ..,255}. Besides, we have a class set for texture pattern, $\Omega = \{\omega_{S1}, \omega_{S2}, ..., \omega_{SP}\}$ where $\omega_{Si} \in M$ and M is the set of available classes. In the proposed system, we have only two classes, the edge and the non-edge classes.

In Markov chain analysis, the conditional probability of certain pixel being certain class is given by Gibbs distribution according to Hammersley-Clifford

theorem. The density function is $\pi(\omega) = \frac{1}{\sum_{\omega} exp(\frac{-U(\omega)}{T})} exp(\frac{-U(\omega)}{T})$, where T is the temperature constant, which is used in stimulated annealing. The energy term can be further represented as $U(\omega, x_i) = V_1(\omega, x_i) + \sum_{i' \in N_i} \beta_{i,i'} \delta(x_i, x_{i'})$, where $V_1(\omega, x_i)$ represents the potential for pixel with certain intensity value belongs to certain class and the $\delta(x_i, x_{i'})$ is the normalised correlation between pixel at s_i and those at $s_{i'}$.

When the texture is being learnt by the feature learning module, the set of $\beta_{i,i'}$ is estimated according to the requirement that the probability of its associated texture class will be maximised. The estimation algorithm used in the system is simulated annealing. The set of $\beta_{i,i'}$ corresponds to the correlation value and thus represents the configuration of the pixels such that it can be classified as that texture class. In the system, this set of estimated β will be used as texture feature vector. It will be used as input of support vector machine such that the association between texture feature and texture class can be formed.

3.2 Texture Learning Using Support Vector Machine

Support vector machine have been widely used in recognition recently due to its non-linear classification power and thus be used to solve complicated recognition problem such as face recognition (e.g. [14]). Given data set: $\{(b_1, y_1), (b_2, y_2), ..., (b_l, y_l)\} \in B \times \{+1, -1\}$, support vector machine can learn to find out the association between b_i and y_i. In the proposed system, the b_i will be the texture feature set $\{\beta_{i,i'}\}$ after texture extraction on the input image and $\{+1, -1\}$ refers to edge and non-edge classes. During learning phase, the support vector machine will be trained. The classifier's parameters, α_i are learnt from data set $\{b_i, y_i\}$ under the criteria function, $max_{\alpha} \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j k(b_i, b_j)$. Gradient ascent approach is used in the system. During testing phase, the texture feature extracted from the image will be classified by the support vector machine. The determinant function can be written as $f(b) = sgn(\sum_{i=1}^{l} \alpha_i y_i k(b, b_i) + constant)$, where $k(\cdot, \cdot)$ is gaussian RBF kernel. The output will be an binary image with '1' indicates the edge class and '0' indicates the non-edge class. Mathematical details of support vector machine can be found in [10].

3.3 Texture Segmentation by Snake Fitting

Active contour [11] had been used in pattern location and tracking [15] for a long time. It is good at attaching to object with strong edge and irregular shape. The snake can be interpreted as parametric curve $v(s) = [x(s), y(s)]$.

In the proposed system, the initial position of the active contour is defined by the user. The active contour will move according to the refined energy function, $E^*_{snake} = \int_0^1 \{[E_{int}(v(s))] + [E_{texture}(v(s))] + [E_{con}(v(s))]\} ds$, where E_{int} represents the internal energy of the snake due to bending, $E_{texture}$ represents the texture-based image forces, and E_{con} represents the external constraint forces. The snake is said to be fitted if the E^*_{snake} is minimised.

The above equation is similar to commonly used snake equation but with the energy term $E_{texture}(v(s))$ replaces the original $E_{image}(v(s))$ which means image force. The energy term $E_{texture}(v(s))$ represents the energy of texture and is proportional to the negative of similarity of desired texture. This means the energy will be lower near the patch that shows desired texture (e.g. edge texture). Thus, the snake will search for strong edge in the binary texture map, that described in Section 3.2, along the direction toward the centroid of potential region. It stops at the pixel with strong edge characteristic in the texture map. Thus, the term $E_{texture}(v(s))$ can be interpreted as the texture attractive force and the snake is texture sensitive. Texture represents a patch of pixels instead of a single pixel and texture-based analysis is more tolerant to noise compare with pixel-based analysis. Thus, texture is a much reliable feature than strong edge under pixel-based analysis.

3.4 Prediction by Kalman Filtering

Kalman filtering [12] is a prediction-correction procedure. By encapsulating the motion of the object into internal states, Kalman filtering aims at finding appropriate states that gives best-fit observations. Dynamic equation and measurement equation will be used in Kalman filter for representing the change in internal states and conversion from internal state to observation respectively.

Since the motion of the vertebrae is planar, kalman filter is good enough to make prediction on the spinal motion. In the system, dynamic model is assumed to be common rigid body motion and the uni-modal Gaussian noise distribution is assumed. The state transition equation and the measurement equation used are: $\begin{bmatrix} x_{t+1} \\ y_{t+1} \\ u_{t+1} \\ v_{t+1} \\ a_{t+1} \end{bmatrix} = \begin{bmatrix} 1\,0\,1\,0\,0 \\ 0\,1\,0\,1\,0 \\ 0\,0\,1\,0\,1 \\ 0\,0\,0\,1\,1 \\ 0\,0\,0\,0\,1 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \\ u_t \\ v_t \\ a_t \end{bmatrix} + m_t$, and $\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} 1\,0\,0\,0\,0 \\ 0\,1\,0\,0\,0 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \\ u_t \\ v_t \\ a_t \end{bmatrix} + n_t$ respectively, where x_t, y_t is the position of the landmark, u_t, v_t represent the velocity along x and y direction respectively, a_t represents the acceleration, and m_t, n_t represent the dynamic noise and measurement noise respectively. The equation for updating the parameters in Kalman filter can be found in [12].

4 Experiment and Result

The proposed system was implemented using Visual C++ under Microsoft Windows. The experiments were done on a P4 2.26 GHz computer with 512M Ram running Microsoft Windows.

4.1 Experiment 1: Low Contrast and Noisy Video Fluoroscopic Image

In this experiment, the performance of the proposed feature learning and detection algorithm was evaluated. The vertebrae had to be segmented from the

medical image with poor quality and low contrast. Actually, the image may not be segmented easily manually. There are totally 2 phases in this experiment. The first phase is learning phase where the texture pattern associated with edges is learnt and the second phase is vertebral boundary detection where the snake is fitted toward the edge detected.

In training phase, the total number of samples to be trained was around 1000, half of them was edge and half of them was non-edge. The samples were selected manually and were selected from images with similar illumination and contrast. The learning images and the testing images were randomly selected from the same video sequence and thus with similar illumination and contrast. In testing phase, the snake is initially mounted at location close to the vertebral boundary. The snake will then attach to the boundary automatically by using texture as heuristic.

The result of segmentation is showed in figure 2. It shows that the snake can fit the target vertebrae very well. The accuracy cannot be determined here due to no ground truth image provided. If the output is compare with the landmarks marked by a skilled physician, the relative root-mean-square error (the difference between the tracked corners and the physician-marked corners) is less than 3% in average out of 100 testing samples. The processing time is around 18s when the whole texture binary map is formed and the image with size 600 x 450 pixels. Some of the edge detection result of commonly used edge detector is shown in figure 3 for reference. It shows that the proposed method works much better than the commonly used edge detectors.

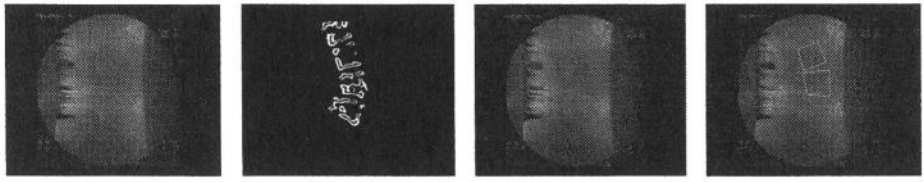


Fig. 2. *The first image shows the testing image. The second image shows the binary image after final classification. The third image shows the fused image constructed from the testing image and the binary image. The fourth image shows the snake attached to the boundary of the vertebrae.*

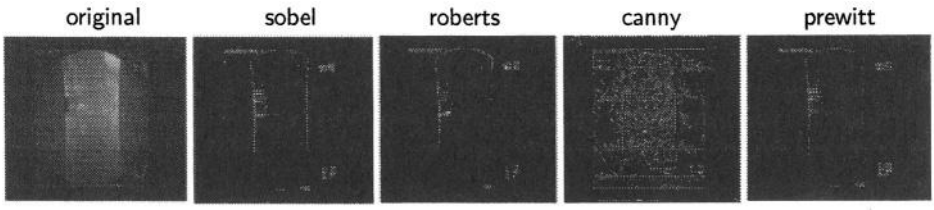


Fig. 3. *Edge detection result of some commonly used edge detector is shown.*

4.2 Experiment 2: Tracking Spine in Videofluoroscopic Video

In this experiment, the performance of the whole system is evaluated. The system ran in the same way as described in section 2. Firstly, the accuracy on the reported location of the vertebra was tested. One of the vertebra in the video was tracked. The corners will be extracted and reported. Throughout a video sequence of 2 minutes, 200 sample frames was tested. The testing result is shown in figure 4(a). It shows that the reported location by the system and the marked location by the physician are very close. The relative root-mean-square error (the difference between the tracked location and the physician-marked location) is less than 5% in average. The processing time of each frame is aound 0.5 s because the edge pattern is now analysed along the snake instead of analysed the whole image.

The accuracy on intervertebral relation reported by the system were also tested. The angle between two vertebrae is usually used in most spinal motion analysis. Thus, the accuracy on the angle reported by the system were evaluated. The measurement methodology is shown in figure 5(a). The testing result is shown in figure 5(b). It shows the relative root-mean-square error (the angular difference between the tracked result and the physician-reported result) is quite large during initial phase but getting smaller after 30 frames. The relative root-mean-square error is lower than 10% in average in later stage.

Finally, the number of vertebrae that can be tracked by the system is evaluated. The result is shown in figure 4(b). It shows that totally four of the vertebrae, namely L2, L3, L4 and L5 can be tracked, provided that the illumination and the contrast is not varied a lot. The relative root-mean-square error reported is less than 10% in these four vertebrae.

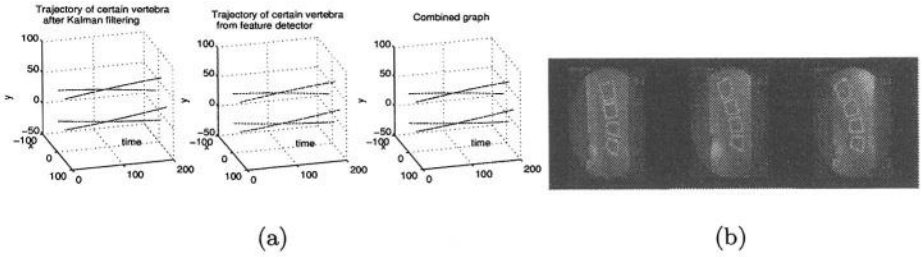


Fig. 4. *(a) The first two graphs show the reported location of the four vertebral corners along the time domain and time series data of location of the corresponding vertebral corners marked by physician respectively. The third graph combines the above two time series into one graph, (b) The tracking result of L2 to L5 vertebrae.*

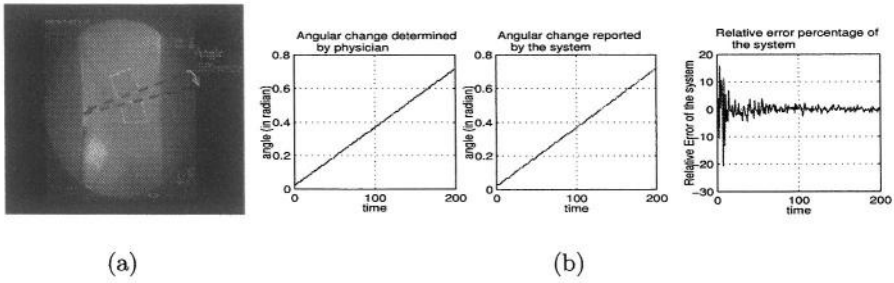


Fig. 5. *(a) The angle difference between middle two vertebrae is recorded in the experiment. (b) The left two graphs show the time series data on angle difference reported by the system and those measured by physician respectively. The third graph shows the corresponding relative root-mean-square error in percentage.*

5 Conclusions

In this paper, a system for automatic spinal motion analysis is proposed. The proposed system requires less human intervention than common approaches by automating the edge detection and snake fitting. Operators may need to setup initial snake position in the first frame only. The edge will then be detected automatically using pattern recognition and the snake will fit toward the edge accordingly. The initial snake position in the next frame will be predicted through the use of dynamic that learnt from previous observations. Experimental results show that the proposed system can segment vertebrae from videofluoroscopic images automatically and accurately.

References

[1] institute for Work & Health: Iwh fact sheet low-back pain. Internet resource, (Avaliable: http://www.iwh.on.ca/media/lowbackpain.php)
[2] clinical Standards Advisory Group: Epidemiology review: the epidemiology and cost of back pain. HMSO. London (1994)
[3] Breen, A.C., Allen, R., Morris, A.: A digital videofluoroscopic technique for spine kinematics. Journal of Medical Engineering and Technology **13** (1989) 109–113
[4] Gatton, M.L., Pearcy, M.J.: Kinematics and movement sequencing during flexion of the lumbar spine. Clinical Biomechanics **14** (1999) 376–383
[5] Cardan, C., Allen, R.: Measurement of spine motion for diagnosis of mechanical problems. Journal of Simulation Modeling in Medicine **1** (2000) 15–19
[6] Page, W.H., Monteith, W., Whitehead, L.: Dynamic spinal analysis - fact and fiction. Chiropractic journal of Australia **23** (1993) 82–85
[7] Smyth, P.P., Taylor, C.J., Adams, J.E.: Automatic measurement of vertebral shape using active shape models. Image and Vision Computing **15** (1997) 575–581
[8] Zheng, Y., Nixon, M.S., Allen, R.: Automated segmentation of lumbar vertebrae in digital videofluoroscopic images. IEEE Trans. Medicial Imaging **23** (2004) 45–52

[9] Chellappa, R., Jain, A.: Markov Random Fields: Theory and Applications. Academic Press (1993)
[10] Scholkopf, B., Burges, C.J.C., Smola, A.J.: Advances in Kernel Methods - Support Vector Learning. MIT Press, Cambridge, MA (1999)
[11] Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. In: Proc. Int. Conf. on Computer Vision. (1987) 259–268
[12] Kalman, R.E.: A new approach to linear filtering and prediction problems. Trans. ASME - Journal of Basic Engineering **82(Series D)** (1960) 35–35
[13] Krishnamachari, S., Chellappa, R.: Multiresolution gauss-markov random-field models for texture segmentation. IP **6** (1997) 251–267
[14] Fransens, R., DePrins, J., Gool, L.J.V.: Svm-based nonparametric discriminant analysis, an application to face detection. In: ICCV03. (2003) 1289–1296
[15] Blake, A., Isard, M.: Active Contours. Springer (1998)

A New Algorithm Based on Fuzzy Gibbs Random Fields for Image Segmentation*

Gang Yan and Wufan Chen

Key Lab for Medical Image Processing, BME, First Military Medical University, Guangzhou, China
yangang@fimmu.edu.cn, chenwf@fimmu.edu.cn

Abstract. In this paper a new unsupervised segmentation algorithm based on Fuzzy Gibbs Random Field (FGRF) is proposed. This algorithm, named as FGS, can deal with fuzziness and randomness simultaneously. A Classical Gibbs Random Field (CGRF) servers as bridge between prior FGRF and original image. The FGRF is equivalent to CGRF when no fuzziness is considered; therefore, the FGRF is obviously a generalization of the CGRF. The prior FGRF is described in the Potts model, whose parameter is estimated by the maximum pesudolikelihood (MPL) method. The segmentation results are obtained by fuzzifying the image, updating the membership of FGRF based on maximum a posteriori (MAP) criteria, and defuzzifying the image according to maximum membership principle (MMP). Specially, this algorithm can filter the noise effectively. The experiments show that this algorithm is obviously better than CGRF-based methods and conventional FCM methods as well.

1 Introduction

Image segmentation is a key technique in the pattern recognition, computer vision and image analysis. The accurately segmented medical images is very helpful for clinical diagnose and quantitative analysis. Automated segmentation is however very complicated, facing difficulties due to overlapping intensities, anatomical variability in shape, size, and orientation, partial volume effects, as well as noise perturbation, intensity inhomogeneities, and low contrast in images [1]. To overcome those difficulties, there has recently been growing interesting in soft segmentation methods [2]-[4]. The soft segmentation, where each pixel may be classified into classes with a varying degree of membership, is a more natural way. Introducing the fuzzy set theory into the segmentation is the outstanding contribution for soft segmentation algorithms. The algorithm is called soft image segmentation scheme if it is based on fuzzy set. When the fuzziness is eliminated according to some rules, the segmented image is exactly obtained.

Although the FCM algorithms are widely used in image segmentation, there are still some disadvantages. It is assumed that the data are of spatial independence or no

* This work was supported by 973 Program of China (No: 2003CB716101) and Key NNSF of China (No: 30130180).

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 163-170, 2004.

context. Those assumptions are unreasonable. The FCM algorithm hardly deals with noised images. It is indispensable to incorporate the contextual constraint into the algorithm during segmentation. Another, the statistical approaches are increasingly used. Among them, Markov random fields (MRF)-based methods are of most importance due to well modeling the prior information [5], [6], but they poorly deal with fuzziness. Furthermore, only hard segmentation was obtained with these methods.

The fuzzy-based methods and MRF-based methods have their respective advantages. It can be predicated that integrating the fuzzy set theory with MRF theory will create wonderful results. The mixing of Markov and fuzzy approaches is discussed in [7]-[9]. Only two-class segmentation is discussed by adding a fuzzy class [7], [8]. H. Caillol and F. Salzenstein only had discussion about generalization to multi-class segmentation. S. Ruan et al used the fuzzy Markov Random Field (FMRF) as a prior to segmented MR medical image [9], which is a multi-class problem. However, only two-tissue mixtures are considered. Three-tissue or more-tissue mixtures are not concerned. The idea of merging the fuzziness and randomness is to be refreshed. The new concept of FMRF based on fuzzy random variable should be proposed. Every pixel is considered as fuzzy case, and is the mixture of all the classes.

The paper is organized as follows. In section 2 some preliminaries about our model are mentioned. The concept of FGRF is represented in Section 3. Our model based on FGRF is described in section 4. Section 5 gives the algorithm and some experiments. The final section is concerning conclusions and discussion on this paper.

2 Preliminaries

Fuzzy set theory is the extension of conventional set theory. It was introduced by Prof. Lotfi A. Zadeh of UC/Berkeley in 1965 to model the vagueness and ambiguity. Given a set U, a fuzzy set $\tilde{A}$ in U is a mapping from U into interval [0,1], i.e.,

$$\tilde{A}: U \to [0,1], x \mapsto \tilde{A}(x),$$

where $\tilde{A}(x)$ is called membership function. Given $x_0 \in X$, $\tilde{A}(x_0)$ is the membership value for element x_0. All fuzzy sets in U are denoted by $\mathbb{F}(U)$.

Fuzzy set $\tilde{A}$ in U is denoted by $\tilde{A} = \{(x, \tilde{A}(x) \mid x \in U\}$. If the set U is finite, then fuzzy set $\tilde{A}$ can be written as $\tilde{A} = \sum \tilde{A}(x_i)/x_i$ or as a fuzzy vector $\tilde{A} = (\tilde{A}(x_1), \tilde{A}(x_2), \cdots, \tilde{A}(x_n))$. We always consider that the fuzzy set is the same notation as its membership function.

Uncertainty of data comes from fuzziness and randomness modeled by the random variables and the fuzzy sets respectively. Fuzzy random variable can model the two kinds of uncertainty first proposed by Kwakernaak [10]. In the definition, fuzzy random variable is a fuzzy-valued mapping. It is necessary to define the probability of fuzzy event, thereby.

The probability of fuzzy events $\tilde{A}: \Omega \to [0,1] \quad \omega \mapsto \tilde{A}(\omega)$ in the probability space $(\Omega, \mathbb{F}, P)$ is defined as

$$P(\tilde{A}) = E(\tilde{A}(\omega)) = \int_{\Omega} \tilde{A}(\omega) P(d\omega),$$

Where E denotes the expectation. If the sample space is the discrete set, then

$$P(\tilde{A}) = \sum_{i=1}^{n} \tilde{A}(\omega_i) p_i \ , \ p_i = P(\omega_i) \ .$$

Generally p_i is called primary probability.

3 Fuzzy Markov Random Fields

Traditional MRF-based segmentation algorithm requires modeling two random fields. For $S = \{1,2,\cdots,n\}$, the set of pixels, $X = (X_s)_{s\in S}$ is unobservable MRF, also called the label field. The image to be segmented is a realization of the observed random field $Y = (Y_s)_{s\in S}$. Random variable must be generalized to fuzzy random variables for treating the vagueness. In this case, each pixel corresponds to a fuzzy set of label field $L = \{1,2,\cdots,k\}$. Soft segmentation can be realized, and the final result may be obtained flexibly by many defuzzification methods.

In detail, each pixel i is attached with a fuzzy set represented by a k-dimensional fuzzy vector $(\mu_{i1},\mu_{i2,}\cdots,\mu_{ik})$, where μ_{is} is the membership value of the ith pixel to the sth class, and the constraint $\sum_{s=1}^{k}\mu_{is} = 1$ is introduced. $X = (X_s)_{s\in S}$ is called fuzzy random field if each X_s is fuzzy random variable. Fuzzy random field is just FMRF if the family of fuzzy random variables is constrained by Markovianity.

When no fuzziness is considered, $\mathbb{F}(L) = \{I_i(l) \mid i \in L\}$ is composed of all the indicator functions of each label, i.e., $\mathbb{F}(L)$ includes no fuzzy cases. The fuzzy random variables will certainly degenerate into classic random variables without fuzziness.

It is essential to know the joint probability of fuzzy event

$$X = x = (X_1 = x_1, X_2 = x_2, \cdots, X_s = x_s) ,$$

where each x_i is fuzzy set. The probability $P(X = x)$ is given in a Gibbs form [6]

$$p(x) = Z^{-1} e^{-U(x)} ,$$

where $U(x)$ stands for the energy function and Z is the normalizing constant.

The family of fuzzy random variables is said to be a fuzzy Gibbs random field (FGRF) if and only if its joint probability obeys the Gibbs form.

4 FGRF Model

When the image segmentation is formulated in a Bayesian rule, the goal is to estimate prior X from a posteriori $P(X \mid Y)$. The prior and the likelihood distribution are presented respectively. We adopt the MAP estimation as the statistical criterion.

The FGRF is used as the prior to describe the context information. A fuzzy Potts model is to be established based on the classical Potts model. Usually the neighborhood system V is the spatial 8-connexity. In fuzzy Potts model, only pair-site clique potential is considered and defined as

$$V_2(x_i, x_j) = \frac{\beta}{2}(\| x_i - x_j \|), \tag{1}$$

where β is parameter, x_i and x_j is the fuzzy set in its labels. The distance between the two fuzzy sets is measured by the hamming distance

$$\| x_i - x_j \| = \sum_{m=1}^{k} | \mu_{im} - \mu_{jm} | .$$

The larger the hamming distance, the larger the difference between two neighboring sites is, and the larger the pair-site clique potential is too.

The subtle difference between two neighboring sites is taken into account in FGRF. It is concluded that FGRF is more powerful and flexible than CGRF. When no fuzziness is considered, we yield the classical Potts model denoted by $Z = (Z_s)_{s \in S}$; the fuzzy Potts model and the classical Potts model are a pair of random field that hold the same parameter thereby.

It is necessary to define the membership function. The membership function is developed using geometrical features [11]. Let N_j be the number of neighborhood pixels of the candidate pixel X_s belonging to class j. The membership function is defined as

$$\tilde{A}_s : L \to [0,1] \quad j \mapsto \tilde{A}_s(j) = N_j / \sum_j N_j \tag{2}$$

where $\tilde{A}_s(j)$ denotes the degree of this pixel belonging to class j. We assume that each class obeys normal form with the mean and variance

$$\theta_y = (m_1, \sigma_1^2; m_2, \sigma_2^2; \cdots; m_k, \sigma_k^2).$$

If the fuzzy random variable X_s takes a value $x_s = (\mu_{s1}, \mu_{s2}, \cdots, \mu_{sk})$. Then the distribution of Y_s conditional on X_s also obeys normal form with mean and variance

$$m(s) = \mu_{s1} m_1 + \mu_{s2} m_2 + \cdots + \mu_{sk} m_k, \ \sigma^2(s) = \mu_{s1}^2 \sigma_1^2 + \mu_{s2}^2 \sigma_2^2 + \cdots + \mu_{sk}^2 \sigma_k^2 .$$

To calculate the likelihood distribution $P(Y \mid X)$, we made the assumptions that the observation components Y_s are conditionally independent given X, and the distribution of each Y_s conditional on X is equivalent to its distribution conditional on X_s. Thus

$$P(Y \mid X) = \prod_i P(Y_i \mid X_i) .$$

The parameter is estimated in MPL method. The concavity of pseudolikelihood function determines simple and fast computation. It is key to deduce a formula so that MPL estimation can be implemented.

For Potts model, its distribution is

$$p(x,\beta) = z(\beta)\cdot e^{-U(x,\beta)},$$

where

$$u(x,\beta) = \sum_{i,j\in N_i} V_2(i,j),\ V_2(i,j) = \begin{cases} 0 & x_i = x_i \\ \beta & x_i \neq x_i \end{cases},\ z(\beta) = \sum_x \exp[-u(x,\beta)].$$

Let $n(i,x_i)$ be the numbers of label different from the candidate pixel x_i in its neighbor. The MPL estimation of parameter β is obtained by solving the following equation

$$t(x) = \sum_{i\in S} E_p[n(i,x_i)] \tag{3}$$

where $E_p[n(i,x_i)]$ is the expectation with respect to conditional distribution $P(x_i \mid x_{N_i}), i \in S$ and $t(x) = \sum_{i\in S} n(i,x_i)$. On the other hand, it is the concavity of the pseudolikelihood function that MPL estimation is assured to be unique [12].

5 Algorithm and Experiments

When the MAP is adopted as the segmentation criteria, the goal is to find X that maximize the posteriori, i.e., minimizing the posterior energy $E(X \mid Y)$

$$E(X \mid Y) = \sum_{i\in S}\left[\frac{(y_i - m_i)^2}{\sigma_i^2} + \log(\sigma_i)\right] + \beta t(x) \tag{4}$$

where $t(x)$, m_i and σ_i^2 were discussed in section 4. The parameter β was estimated in pseudolikelihood method using Eq.(3). The unknown parameters is denoted by

$$\theta_d = \{(m(i),\sigma^2(i)) \mid i = 1,2,\cdots,k\}.$$

The FGRF is denoted by the $X = (X_s)_{s\in S}$ whose no fuzzy case is CGRF $Z = (Z_s)_{s\in S}$. It is easy to understand that the CGRF builds a bridge between the prior FGRF and the observed image. In detail, our algorithm is described in the follow steps:

1. Set the class number. The initial segmentation $Z^{(0)}$ is obtained in the k-means clustering procedure, and then the parameters $\theta_d^{(0)}$ can also be initialized;
2. Estimate the parameter β and update during each iteration.
3. Fuzzify the CGRF $Z^{(0)}$ and obtain the initial value of FGRF $X^{(0)}$;
4. Update FGRF X using iterated conditional modes (ICM) by solving Eq. (5)

$$x^{(k+1)} = \arg\max_{x^{(k)}} \sum_{i \in S} \left[\frac{(y_i - \mu_i^{(k)})^2}{(\sigma_i^{(k)})^2} + \log(\sigma_i^{(k)}) \right] + \beta t(x^{(k)}) \tag{5}$$

5. Defuzzify the $x^{(k+1)}$ using MMP and yield an updated CGRF $z^{(k+1)}$.
6. Update the parameter θ_d using the empirical means and variances.
7. Repeat the step 4)-6) until convergence.

Our algorithm is tested on both simulated MR images from the Brain Web Simulated Brain Database at the McConnell Brain Imaging Center of the Montreal Neurological Institute (MNI), McGill University, and real MR images. Simulated brain images are corrupted with different noise level. The control algorithms are the classical GRF (CGS), maximum likelihood (ML) and fuzzy c-mean (FCM) algorithm.

Fig. 1 presents a comparison results. For FGS algorithm, the original image has been divided into the distinct and exclusive regions. Moreover it has more advantages than the control algorithms, such as smooth and continuous boundary and no noise. Fig. 2 present a comparison of segmentation results for image corrupted with 9% noise level. It is nearly no ability for FCM or ML algorithms to filter the noise. It is doubtless that FGS algorithm is much more powerful than CGS algorithm for filtering noise.

To further testify the powerful property in filtering the noise, our method is also realized to segment a simulated image corrupted greatly with unknown noise level. Fig. 3 shows that only our algorithm can obtain the correct segmentation.

To measure the robustness of the algorithms, the overlap metric is utilized as the criteria. The overlap metric is a measure for comparing two segmentations that is defined for a given class assignment as the sum of the number of pixels that both have the class assignment in each segmentation divided by the sum of pixels where either segmentation has the class assignment [13]. Larger metric means more similar for results. The segmented images corrupted by different noise level are compared with the no noise image using the different algorithms. Experiments show that our algorithm is much more robust than the others. Table 1 gives the overlap metrics of white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF). It is satisfied that the overlap metric of our algorithm varies slowly with the noise level increasing, i.e., our algorithm is insensitive to noise.

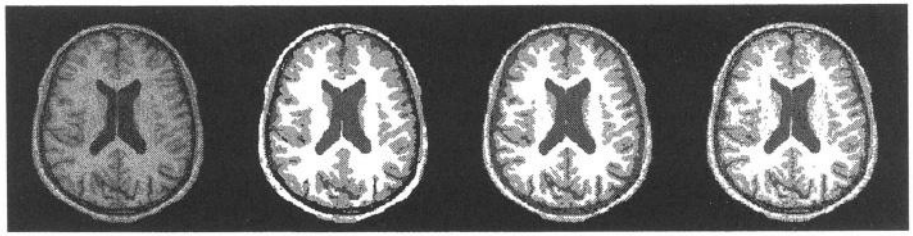


Fig. 1. Comparison of segmentation results on real clinical MR image. (a) Original image, (b) using FGS, (c) using CGS, (d) using FCM

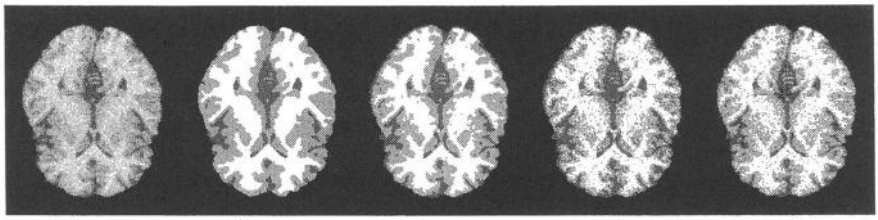


Fig. 2. Comparison of segmentation results on stimulated MR image, (a) Original images with 9% noise level, (b) using FGS, (c) using CGS, (d) using ML, (e) using FCM

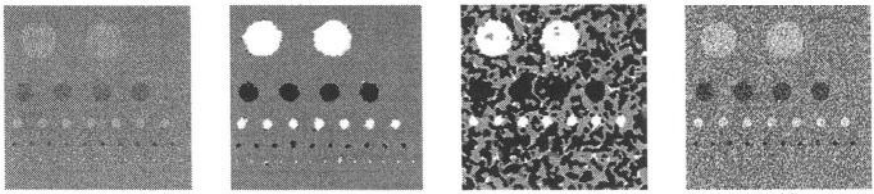


Fig. 3. Comparison of segmentation results on general stimulated image. (a) Original image, (b) using FGS, (c) using CGS, (d) using FCM

Table 1. Overlap metric with different noise level

Noise level	Overlap metric of WM				Overlap metric of GM			
	FGS	FCM	CGS	ML	FGS	FCM	CGS	ML
1%	0.97	0.98	0.97	0.88	0.96	0.97	0.96	0.87
3%	0.94	0.94	0.92	0.73	0.93	0.92	0.87	0.77
5%	0.93	0.89	0.91	0.69	0.90	0.86	0.87	0.72
7%	0.91	0.83	0.89	0.65	0.86	0.79	0.85	0.65
9%	0.90	0.76	0.85	0.61	0.85	0.70	0.81	0.58

6 Conclusion and Discussion

The proposed algorithm takes into account the fuzziness and the randomness simultaneously. Each pixel is modeled by fuzzy random variable. The FGRF is used to obtain the contextual information. All the experiments show that our algorithm can obtain accurate segmentations. But the intensity inhomogeneity is not taken into account. We will try to settle this problem in the following work.

Reference

1. J.-H. Xue, S. Ruan, B. Moretti, M. Revenu, and D. Bloyet: Knowledge-based segmentation and labeling of brain structures from MRI images. Pattern Recognition Letters, Vol. 22, No. 3-4, pp. 395-405, 2001
2. W.C. Liew, H. Yan: An Adaptive Spatial Fuzzy Clustering Algorithm for 3-D MR Image Segmentation. Vol. 22, No. 9, pp. 1063-1075, 2003
3. M. N. Ahmed, S. M. Yamany: A Modified Fuzzy C-Means Algorithms for Bias Field Estimation and Segmentation of MRI Data. IEEE Trans. Med. Imag., Vol. 21, No. 3, pp. 193-199, 2002
4. C. Z. Zhu, T. Z. Jiang: MultiContext fuzzy clustering for separation of brain tissues in magnetic resonance images. NeuroImage, Vol.18, No.3, pp. 685-696, 2003
5. Y. Zhang, M. Brady, S. Smith: Segmentation of Brain Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm. IEEE Trans. Med. Imag., Vol. 20, No.1, pp. 45-57, 2001
6. Ciro, D'Elia, G. Poggi, and G. Scarpa: A tree-Structured Markov Random Field Model for Bayesian Image Segmentation. IEEE Trans. On Image Processing, Vol.12, No. 10, pp. 1259-1273, 2003
7. H. Caillol, W. Pieczynski, and A. Hillion: Estimation of Fuzzy Gaussian Mixture and Unsupervised Statistical Image Segmentation. IEEE Trans. On Image Processing, Vol. 6, No. 3, pp. 425-440, 1997
8. F. Salzenstein and W. Pieczynski: Parameter estimation in hidden fuzzy Markov random fields and image segmentation. Graphical Models and Image Processing, Vol. 59, No. 4, pp. 205-220, 1997
9. S. Ruan, B. Moretti, J. Fadili and D. Bloyet: Fuzzy Markovian Segmentation in Application of Magnetic Resonance Images. Computer Vision and Image Understanding, Vol. 85, No. 1, pp. 54-69, 2002
10. H.Kwakernaak: Fuzzy random variables-I, Definition and Theorems. Information Sciences, 15, pp. 1-29, 1978
11. G. C. Karmakar and L. S. Dooley: A generic fuzzy rule based image segmentation algorithm. Pattern Recognition Letters 23, pp. 1215-1227, 2002
12. Gerhard Winkler: Image Analysis, Random Fields and Dynamic Monte Carlo Methods. Heidelberg, Germany: Springer-Verlag, 1995
13. Available: http://www.bic.mni.mcgill.ca/brainweb.

Improved Fiber Tracking for Diffusion Tensor MRI

Mei Bai, Shuqian Luo

Capital University of Medical Sciences, Beijing 100054,China
baimei12@yahoo.com

Abstract. Diffusion tensor MRI is currently the only imaging method that can provide information about water molecule diffusion direction, which reflects the patterns of white matter connectivity in the human brain. This paper presents a fiber-tracking algorithm based on an improved streamline tracking technique (STT). Synthetic datasets were designed to test the stability of the fiber tracking method and its ability to handle areas with uncertainty or isotropic tensors. In vivo DT-MRI data of the human brain has also been used to evaluate the performance of the improved STT algorithm, demonstrating the strength of the proposed technique.

1 Introduction

Diffusion tensor MRI (DT-MRI) is an in vivo imaging modality with the potential of generating fiber trajectories of the human brain to reflect the anatomical connectivity. Furthermore, the various water diffusion information provided by DT-MRI can reflect microstructure and texture characteristics of the brain tissues [1,2,3]. Thus far, there is no "gold standard" for in vivo fiber tractography [4]. In vitro validation of the fiber tract obtained by DT-MRI has been attempted histologically [5,6], but sample dissection, freezing, dehydration, and fixation can potentially change the microstructure of the tissue and distort the histological sample. Significant advances have been achieved in recent years by using the tract-tracing methods based on chemical tracers, in which chemicals agents are injected and their destinations are confirmed [7]. Due to the nature of the experiment design, these techniques are not suitable for in vivo studies. Quantitative validation of the virtual fiber bundles obtained by DT-MRI is an important field of research. The purpose of this paper is to describe an improved STT algorithm for fiber tracking, which is validated with both synthetic and in vivo data sets.

2 Material and Methods

2.1 Synthetic Fiber Models

In general, diffusion tensor can be visualized by using an ellipsoid where the principal axes correspond to the directions of the eigenvector system [8]. Let $\lambda_1 \geq \lambda_2 \geq \lambda_3$ be the

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 171-178, 2004.

eigenvalues of the symmetric diffusion tensor D, and $\hat{e}_i$ be the eigenvector corresponding to λ_i. The tensor D can be described by

$$D = \lambda_1 \hat{e}_1 \hat{e}_1^T + \lambda_2 \hat{e}_2 \hat{e}_2^T + \lambda_3 \hat{e}_3 \hat{e}_3^T \quad (1)$$

If the tensor of a particular point is isotropic, the corresponding tensor D can be described as follows

$$D_{ref} = D + \hat{e}_1 \hat{e}_1^T \quad (2)$$

The isotropic tensor can be changed into anisotropic tensor, as shown in Fig.l. By changing the direction of $\hat{e}_1$ we can control the direction of major eigenvector D_{ref}.

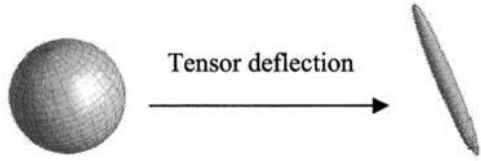


Fig. 1. Isotropic tensor transforms to anisotropic tensor using tensor deflection

In this study, three different synthetic fiber orientation data sets are used. The first simulated tensor field is parameterized as a group of parabola:

$$x = 0.05y^2 + a \quad (3)$$

where a is a displacement constant that $1 \le a \le 10$. In this case, the direction of the corresponding eigenvector is the tangent of the group of curves as shown in Fig.2(a), where arrows indicate the major eigenvector direction at each point. In regions of x=1:10, y=1:10, the eigenvector directions change along to the tangent of $x = 0.05y^2 + a$, whereas in regions of x=11:20, y=1:10, the eigenvector directions change along the tangent of $x = -0.05y^2 + a$. Finally, within the area of x= 1:20, y=11:20, the directions of the eigenvectors are all parallel to the y axis.

Fig.2(b) describes another synthetic data set where the eigenvector directions within the region of x=1:9, y=1:20 and x=11:20, y=1:20 change along the tangent of $y = -0.25(x-10)^2 + b$ (b is a displacement constant), and in the region of x=10, y=1:20, the tensor becomes isotropic. The third synthetic data set simulates 3D fiber distribution as shown in Fig. 2(c), where the primary eigenvectors are aligned with part of a helix described by:

$$\begin{aligned} x &= 100 * \sin(t) \\ y &= 100 * \cos(t) \\ z &= 30 * t \end{aligned} \quad (4)$$

In Eq. (4), t is from 0.3π to 0.6π and the data size is 25×96×31 pixels.

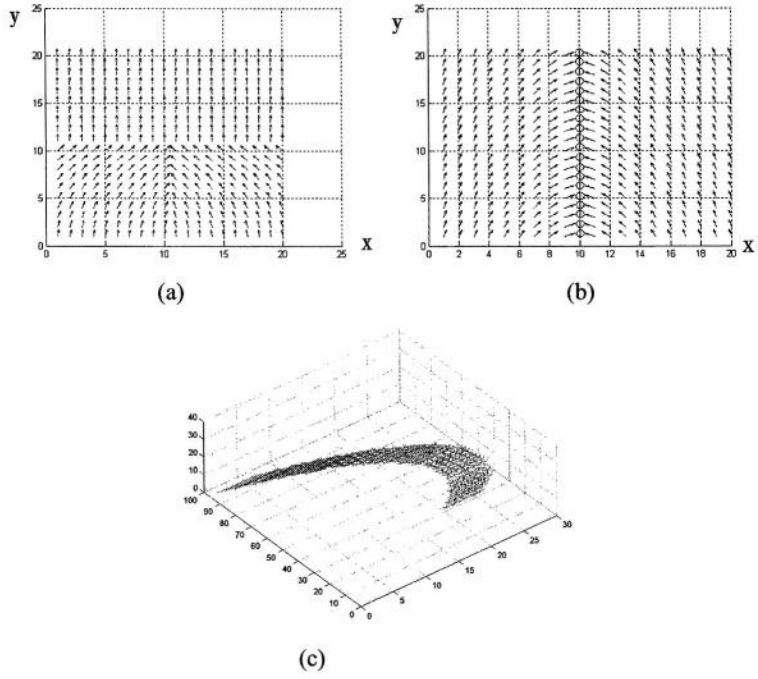


Fig. 2. Simulation data sets used for validating the fiber tracking algorithm, where (a) and (b) are modeled by parabolas and (c) corresponds to part of a helix.

2.2 In Vivo Data

Clinical DT-MRI images were acquired on 1.5T MRI scanner (Sonata, Siemens, Germany). A set of DWIs was obtained by using an echo-planar sequence. Imaging parameters were as follows: FOV= $23\times23\ cm^2$, matrix=128×128, slice thickness=5 *mm,* TR=5*s*, TE=101 *ms,* NEX=4, bvalue=$500\ s/mm^2$. Seven diffusion gradient vectors were as follows: (0,0,0), (1,0,1), (-1,0,1), (0,1,1), (0,1, -1), (1,1,0) and (-1,1,0). Inside tracing ROI, all pixels which FA is larger than 0.2 were selected as seed points.

2.3 Algorithm for Fiber Tracking

Currently there are several approaches for reconstructing fiber tracts. Most of the techniques for tracking nerve fibers based on DT-MRI are variants of the Principal Diffusion Direction (PDD) Tracking algorithm. The main idea of PDD is to start at a given starting point x_0 and proceed in the direction of the main eigenvector, i.e., the PDD $\hat{e}_1(x_t)$:

$$\dot{x}_t = \hat{e}_1(x_t) \tag{5}$$

$$x_{t+1} = x_t + \alpha \dot{x}_t \tag{6}$$

The tracking process is stopped when the anisotropy of the diffusion tensor reaches a lower threshold or when the tracking direction starts to change abruptly. The tensor at position x_{t+1} could be interpolated from tensor field D.

Basser et al. [9] introduced STT for brain fiber tracking by assuming that the major eigenvector direction is tangent to the tract pathway. They assume that a white matter fiber tract could be represented as a 3D space curve, i.e., a vector $\mathbf{r}(s)$, parameterized by the arc length, s, of the trajectory shown in Fig. 3.

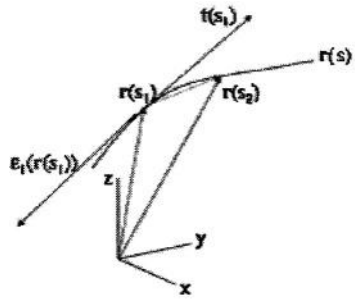


Fig. 3. Representation of a white fiber trajectory

Let $\mathbf{t}(s)$ be the unit tangent vector at position $\mathbf{r}(s)$, then:

$$\frac{d\mathbf{r}(s)}{ds} = \mathbf{t}(s) \tag{7}$$

The basic method of STT is to equalize the tangent vector $\mathbf{t}(s)$ and the unit eigenvector ε_1, calculated at position $\mathbf{r}(s)$ thus

$$\mathbf{t}(s) = \varepsilon_1(\mathbf{r}(s)) \tag{8}$$

and

$$\frac{d\mathbf{r}(s)}{ds} = \varepsilon_1(\mathbf{r}(s)) \tag{9}$$

The initial condition of above differential equation is $\mathbf{r}(0) = \mathbf{r}_0$, which specifies a starting point on the fiber tract. This differential equation can be resolved by using Euler's method:

$$x_{t+1} = x_t + h\dot{x}_t \tag{10}$$

where h is a coefficient from 0 to 1.

Due to image noise, the major eigenvector direction of the points on the assumed fiber pathway often deviates from the true fiber direction, resulting in misleading sudden direction changes. This makes STT terminate at positions where it should

continue. Furthermore, when a diffusion tensor has a planar or spherical shape, the major eigenvector direction may not be parallel to the true nerve fiber direction; therefore the uncertainty of fiber tracking using STT may become large. To improve the accuracy of STT tracking, a tensor deflection is added towards the previous tracking direction:

$$D_{reg} = D + \alpha \dot{x}_{t-1} \dot{x}_{t-1}^T \tag{11}$$

Therefore the final tracking direction becomes:

$$\begin{aligned} \dot{x}_t &= \hat{e}_1(D_{reg}) \\ &= \hat{e}_1(D + \alpha \dot{x}_{t-1} \dot{x}_{t-1}^T) \end{aligned} \tag{12}$$

By changing the value of α we can adjust the tracking direction. A larger value of α will deflect the proposed fiber paths to a greater extent.

2.4 Continuity of Tracking Direction and Termination Criteria

One of the problems needs to be considered is how to assure the continuity of the tangent vector's direction as described in Eq. (8). In this study, we used the dot product of tangent vectors obtained in the previous step and the one got in the present step. If the result is positive we keep the sign of the current eigenvector; if the result is negative we reverse its sign. For termination, one of the criteria we used is the extent of anisotropy. The fractional anisotropy of the gray matter is typically in the range of 0.1-0.2 [10], by which an empirically value of 0.2 is used as the threshold for termination. Another criterion is the directional change between adjacent pixels. For this purpose, we check the trajectory deflection angles in the tracking process and the algorithm terminates when the angle change becomes too large. The continuity in the fiber orientation can be expressed as:

$$C = \left| \dot{x}_t \cdot \dot{x}_{t-1} \right| \tag{13}$$

3 Experimental Results

Fig. 4 shows the tracking results of the first two synthetic data sets by using the STT and improved STT methods, respectively. The first pixels of each column are seed points used for tracking. Fig. 5 shows the corresponding trajectories of the synthetic 3-D fiber model. All points in the first slice (z=1) are used as the seed points for tracking. It is evident from these figures that STT cannot follow the true fiber tracts and can cause severe deviation from the real fiber orientation. The improved STT, on the other hand, is much more robust that can obtain the whole trajectories.

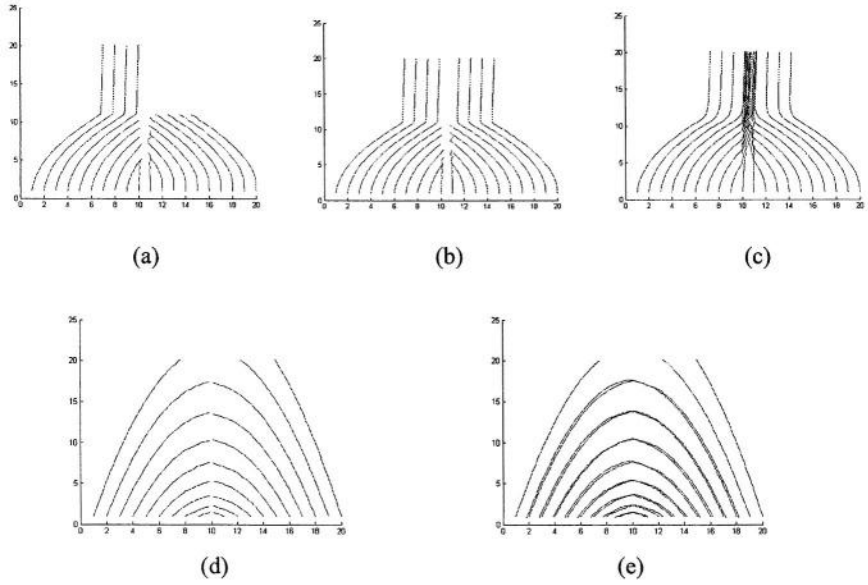


Fig.4. (a)-(c) show tTracking results for the synthetic data set shown in Fig. 2(a) by using STT, and improved STT with $\alpha=0.1$ and $\alpha=0.8$, respectively, (d)-(e) are the racking results for the synthetic data set shown in Fig. 2(b), with STT and improved STT, respectively. For STT, when tracking reaches the area of spherical tensor, the process misses the right direction and is terminated. For the improved STT, the tracking process can handle the isotropic point and keep tracking until the end of the fiber tract.

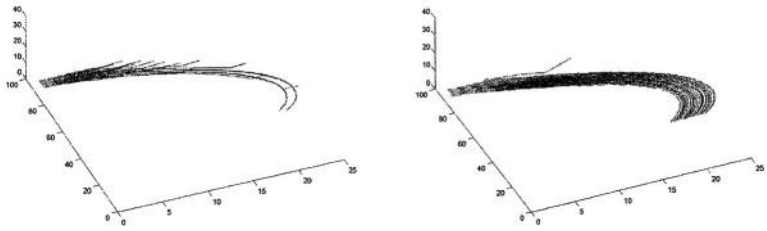


Fig. 5. The tracking result for the 3-D fiber model by using STT (left) and improved STT (right) respectively.

As shown in Fig. 6, nerve fiber tracts inside the ROI for the in vivo data set are almost parallel, without intersection or branching. The tracking results of two methods did not show significant differences.

However, in regions where the distribution of the nerve fiber is complex, e.g. when there are fiber crossing or fiber branching, the STT method is no longer being able to delineate the true fiber path when reaching the "crossing" area. The improved STT can resolve this problem and continue with the tracking process. It is evident that STT is sensitive to noise and has difficulty propagating through spherical and planar

tensors. The modified STT method improves the overall stability of tracking process and significantly decreases the uncertainty of fiber tracking

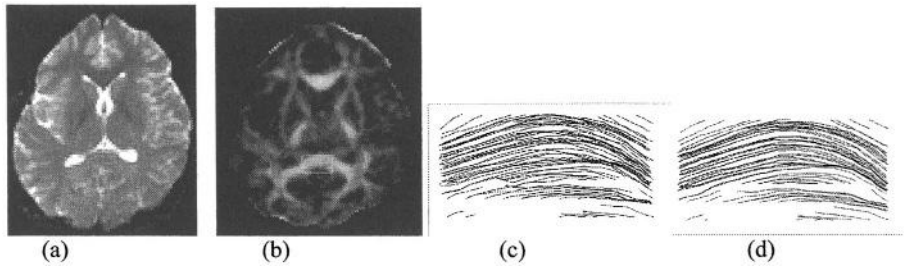


Fig. 6. In vivo tracking results when the fiber shape is relatively simple. (a) shows the selected ROI (in green square), and (b) is the corresponding FA, (c) the tracking result of STT, and (d) is the tracking result of improved STT with $\alpha = 0.2$.

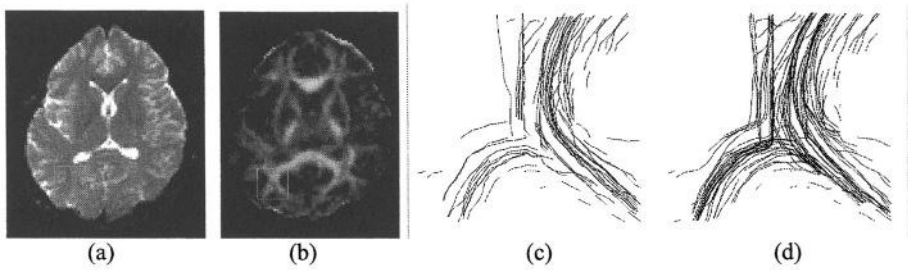


Fig. 7. Tracking results where fiber shape is complex. (a) shows a selected ROI (in green square), and (b) is the corresponding FA, and (c) is the tracking result of STT, (d) improved STT with $\alpha = 0.8$.

4 Discussion and Conclusion

In this paper, we have presented an improved STT method for fiber tracking with DT-MRI. The synthetic tensor field data sets are of practical value for validating the fiber tracking method. It is worth noting that the synthesized tensor fields in this study is relatively simple and further studies are needed to create realistic and more complex fiber models. They should include the consideration of properties such as fiber curvature and connection between fiber orientations. Although the improved STT method can overcome some disadvantages of STT, it still uses PDD as fiber tracking direction, and the problem that tracking result deviates from the real trajectory due to image noise or partial volume effect is unsolved. The issue of how to best map the real fiber tract from DT-MRI voxels should be investigated further.

References

1. Neil J, Miller J, Mukherjee P, Huppi PS. Diffusion tensor imaging of normal and injured developing human brain. NMR Biomed. 2002; 15:543-552.
2. Lori NF, Akbudak E. Diffusion tensor fiber tracking of brain connectivity: Reliability analysis and biological results. NMR Biomed.2002; 15:494-515.
3. Horsfield MA, Jones DK. Applications of diffusion-weighted and diffusion tensor MRI to white matter diseases. NMR Biomed.2002; 15:570-577.
4. Crick F, Jones E. Backwardness of human neuroanatomy. Nature 1993; 361:109-110.
5. Scollan DF, Holmes A, Winslow R, Forder J. Histological validation of myocardial microstructure obtained from diffusion tensor magnetic resonance imaging. Am J Physiol 1998;275:H2308-2318.
6. Hsu EW, Muzikant AL, Matulevivius SA, Penland RC, Henriquez CS. Magnetic resonance myocardial fiber-orientation mapping with direct histological correlation. Am J Physiol 1998; 274:H 1627-1634.
7. Denis LB, Peter van Zijl. From the diffusion coefficient to the diffusion tensor. NMR Biomed. 2002; 15:431-434.
8. Westin, C.-F., Peled, S., Gudbjartsson, H., Kikinis, R., Jolesz, F., 1997. Geometrical diffusion measures for MRI from tensor basis analysis. In: ISMRM '97, Vancouver, Canada, p. 1742.
9. Peter J. Basser, Sinisa Pajevic. In Vivo Fiber Tractography Using DT-MPI Data. Magn.Reson.Med. 2000; 44:625-632.
10. Susumu Mori and Peter C.M.van Zijl. Fiber tracking: principles and strategies – a technical review. NMR Biomed. 2002; 15:468-480.

Rapid and Automatic Extraction of the Modified Talairach Cortical Landmarks from MR Neuroimages

Qingmao Hu, Guoyu Qian, Wieslaw L. Nowinski

Bioinformatics Institute, Singapore 138671
{huqm, qiangy, wieslaw}@bii.a-star.edu.sg

Abstract. An automatic algorithm to locate the modified Talairach cortical landmarks is proposed. Firstly, three planes containing the landmarks are determined, and the optimum thresholds robust to noise and inhomogeneity are calculated based on range-constrained thresholding. Then, the planes are segmented with the chosen thresholds and morphological operations. Finally the segmentation is refined and landmarks are located. The algorithm has been validated against 62 T1-weighted and SPGR MR diversified datasets. For each dataset, it takes less than 2 seconds on Pentium 4 (2.6 GHz) to extract the 6 modified Talairach cortical landmarks. The average landmark location error is below 1 mm. The algorithm is robust and accurate as the factors influencing the determination of cortical landmarks are carefully compensated. A low computational cost results from selecting three 2D planes to process and employing only simple operations. The algorithm is suitable for both research and clinical applications.

1 Introduction

The Talairach transformation ([9]), despite its limitation ([7]), is the most popular way to normalize brains. It is solely determined when the midsagittal plane (MSP), position of the anterior commissure (AC) and posterior commissure (PC), and localization of the 6 Talairach cortical landmarks are available. So far, the Talairach cortical landmarks are determined manually ([3]). Nowinski ([7]) studied the drawbacks of the existing Talairach cortical landmarks and proposed the modified Talairach cortical landmarks. Automatic identification of either the Talairach or modified Talairach cortical landmarks from MR neuroimages is difficult due to the inherent nature of the MR neuroimages: noisy, gray level intensity inhomogeneity, the partial volume effect due to big voxel sizes, sagittal sinus/meninges connected to the cortex, closeness of the cortex to the optic nerves both spatially and in gray levels.

This paper focuses on robust and fast extraction of the 6 modified Talairach cortical landmarks based on anatomic knowledge and range-constrained thresholding. The algorithm has been tested against 62 T1-weighted and SPGR morphological MR datasets, both phantom and real datasets with numerous variations in imaging parameters (noise levels, inhomogeneities, voxel sizes, scanning orientations).

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 179-187, 2004.

2 Material and Method

2.1 Material

62 MR neuroimage datasets collected from four sources have been used for this study including 24 clinical datasets (T1-weighted and SPGR), 20 normal T1-weighted datasets from the Internet Brain Image Segmentation Repository (IBSR) (www.cma.mgh.harvard.edu/ibsr) (some of them are with significant intensity inhomogeneity), and 18 T1-weighted Brain Web phantom datasets (www.bic.mni.mcgill.ca/brainweb) with noise level 0-9%, and inhomogeneity level 0, 20%, and 40%. All data were not corrected by any preprocessing.

2.2 Method

The inputs to our algorithm are MSP ([6]), and the AC and PC ([10]). Fig. 1 shows the flow chart to determine the 6 modified Talairach cortical landmarks.

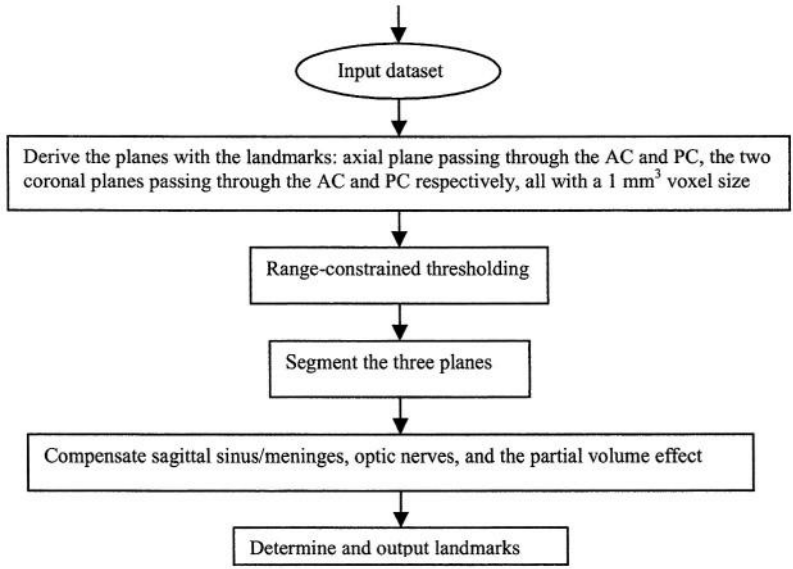


Fig. 1. . Flow chart for determining the modified Talairach cortical landmarks

2.2.1 Range-Constrained Thresholding

The existing thresholding methods ([2], [8]) lack the mechanisms to incorporate knowledge about the images to be segmented and thus handle poorly the inherent nature of the neuroimages like noise and inhomogeneity. We proposed range-constrained thresholding ([5]) which explicitly incorporates the knowledge into the segmentation scheme and consists of 3 steps. The region of interest (ROI) is firstly

determined from the image. Then, within the ROI a range in the corresponding histogram is estimated by knowledge, which represents the maximum and minimum bounds that the background proportion can be. Finally, the threshold is selected to minimize the classification error within this range. Let h(i) denote the frequency of gray level r_i ($0 \le r_i \le 255$), then the accumulative frequency H(i) is $\sum_{r_0}^{r_i} h(i)$, and the frequency at interval $[r_m, r_n]$ is $\sum_{r_m}^{r_n} h(i)$. The following steps will yield the optimum threshold maximizing the between class variance:

1. Specify the two percentages H_l^b and H_h^b, corresponding to the lower and upper frequency bounds of the background in the ROI based on prior knowledge or tests;
2. Calculate r_{low}, which is the gray level corresponding to the background lower bound H_l^b : $r_{low} = \min_i \{i \mid H(i) \ge H_l^b\}$;
3. Calculate r_{high}, which is the gray level corresponding to the background upper bound H_h^b : $r_{high} = \min_i \{i \mid H(i) \ge H_h^b\}$;
4. Calculate the between-class variance with respect to the variable r_k:

$$\Pr(C1) \times D(C1) + \Pr(C2) \times D(C2) \quad (1)$$

where r_k falls within (r_{low}, r_{high}),

$$\Pr(C1) = \sum_{r_{low}}^{r_k} h(i), \quad \Pr(C2) = \sum_{r_k+1}^{r_{high}} h(i), \quad D(C1) = (\mu_0 - \mu_T)^2, \quad D(C2) = (\mu_1 - \mu_T)^2,$$

$$\mu_T = \sum_{r_{low}}^{r_{high}} i \times h(i), \quad \mu_0 = \sum_{r_{low}}^{r_k} i \times h(i), \quad \mu_1 = \sum_{r_k+1}^{r_{high}} i \times h(i).$$

The optimum threshold is the r_k maximizing formula (1) for a specification of H_l^b and H_h^b.

2.2.2 Determination of the A, P, L, and R Landmarks

The AP plane is an axial slice perpendicular to the MSP and passes through the AC and PC. For Talairach transformation, only the u (horizontal) coordinates of the L and R landmarks, and the v (vertical) coordinates of the A and P landmarks are needed.

The following steps will yield the 4 landmarks:

1. Find the voxels enclosed by the skull and take them as the ROI ([1]).
2. Determine the optimum threshold within the ROI, giving $H_l^b = 14\%$, $H_h^b = 28\%$. The two percentages are derived from analyzing the ground-truth segmentation of the 20 IBSR datasets. The threshold maximizing formula (1) is denoted as θ_1.
3. Segment the AP plane by the following sub-steps:
   a) Perform distance transformation of the ROI and convert the distance codes to distance indexes ([4]). Denote the maximum distance index as maxDSkull.
   b) Binarize the image within the ROI and denote the result as BWAP1(u,v).
   c) Perform morphological opening with 3x3 structuring element (SE) with respect to BWAP1(u,v) to get BWAP2(u,v).
   d) Perform morphological opening with 5x5 SE with respect to BWAP1(u,v) to get BWAP3(u,v).

e) Find the connected foreground components of BWAP3(u,v). A foreground component is judged as an object component when its minimum distance index (minD) is bigger than a value (implemented as 20) or maximum distance index (maxD) is bigger than another value (implemented as maxDSkull/2).

f) The object voxels are excluded from the foreground voxels of BWAP2(u,v). Find the connected foreground components of BWAP2(u,v). A foreground component of BWAP2(u,v) is categorized as an object component only when the shape of the component is not similar to meninges. According to anatomical knowledge, meninges have a similar shape to the outline of skull and are quite thin. So when (maxD-minD) is smaller than 0.1 times the number of voxels of this component, the foreground component is judged as background; otherwise it is classified as an object component.

4. Restore object voxels around the object boundaries due to the morphological opening when their gray level is bigger than $\theta 1$.
5. Restore object voxels due to the partial volume effect. The basic idea is to check if the gray level is monotonically decreasing from cortical surface to the background and the cortex proportion of the immediate background voxel is at least 0.5.
6. The v coordinates of the A and P landmarks are the minimum and maximum v coordinates, respectively, of all object voxels. Similarly, the u coordinates of the L and R landmarks are the minimum and maximum u coordinates of all object voxels, respectively. For the AP plane, its u and v coordinates are the same as x and y coordinates, respectively, of the volumetric dataset.

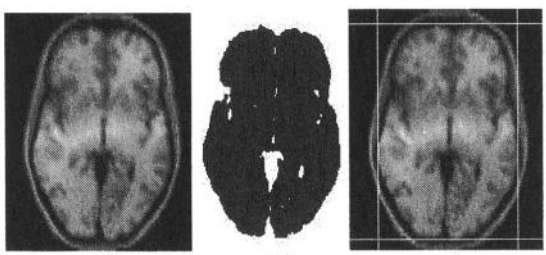


Fig. 2. Identification of the A, P, L and R landmarks from the AP plane: a) the original AP plane; b) the segmented AP plane; and c) the two horizontal lines passing through the v coordinates of the extracted A and P landmarks, and the two vertical lines passing through the u coordinates of the L and R landmarks overlaid on the original AP image

2.2.3 Determination of the S Landmark

The VPC plane is a coronal slice perpendicular to both the MSP and AP plane, and passes through the PC. For the S landmark, its v (vertical) coordinate is the smallest v coordinate of all cortical voxels on the VPC plane.

The S landmark is localized through segmentation of a virtual slice aVPC(u,v) with the complete skull. Denote the VPC plane as VPC(u,v), the coordinates of the PC on VPC(u,v) is (pcU, pcV). aVPC(u,v) is constructed in the following way: aVPC(u,v) is equal to VPC(u,v), when v is not bigger than pcV; aVPC(u,v) equals VPC(u, pcV+pcV-v), when v is bigger than pcV and smaller than (pcV+pcV).

The S landmark is located through segmentation of aVPC(u,v) as follows:

1. Find the ROI of aVPC(u,v). This is the same as finding the ROI of the AP plane.
2. Determine the optimum threshold within the ROI, giving $H_l^b = 20\%, H_h^b = 40\%$. The two percentages are derived from analyzing the ground-truth segmentation of the 20 IBSR datasets. The threshold maximizing formula (1) is denoted as θ_2.
3. Segment the aVPC plane through the same sub-steps as segmenting the AP plane, using threshold θ_2.
4. Restore object voxels around the object boundaries due to the morphological opening as done for the segmentation of the AP plane.
5. Restore object voxels due to the partial volume effect in a similar way as done for the A, P, L, and R landmarks.
6. The minimum v coordinate of all object voxels in aVPC(u,v) is found and is taken as the v coordinate of the S landmark.

Figs. 3 a, b, and c show the derived aVPC, segmented aVPC and the white line to mark the S landmark on the original VPC, respectively.

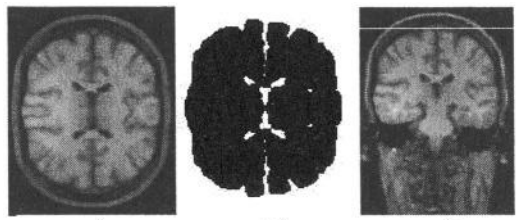


Fig. 3. Identification of the S landmark through segmenting the virtual slice aVPC: a) the derived slice aVPC with complete skull; b) the segmented aVPC slice; and c) the horizontal line passing through the v coordinate of the extracted S landmark overlaid

2.2.4 Determination of the I Landmark

The VAC plane is a coronal slice parallel to the VPC plane and passes through the AC. The v (vertical) coordinate of the I landmark is is the maximum v coordinate of all cortical voxels on the VAC plane.

According to the Talairach-Tournoux atlas ([9]) it can be assumed that the maximum z coordinate difference between the AC and I landmark is within 50 mm.

Denote the AC's coordinates in the VAC(u,v) as (acU, acV). The I landmark is obtained through the following steps:

1. Binarize VAC(u,v) using threshold θ_2 and denote the result as BWVAC1(u,v).
2. Connect the region around the AC to make subsequent seeding feasible.
3. The vertical line passing through the AC divides the VAC into left and right halves. Set the voxels on the vertical line with a bigger v coordinate than (acV+3) mm to background to force the foreground separation in the lower region of BWVAC1(u,v).
4. Perform morphological opening operation with regard to BWVAC1(u,v) with 3x3 SE to get BWVAC2(u,v).

5. Perform morphological erosion with respect to BWVAC2(u,v) with 3x3 SE to get BWVAC3(u,v).
6. Seed from (acU, acV) with respect to BWVAC3(u,v) to get the foreground component. Then, perform morphological dilation on the seeded foreground component with 3x3 SE to get BWVAC4(u,v). The erosion followed by seeding and dilation is intended to break the connection between the cortex and non-cortical structures while preserving the original shape of the cortex.
7. Find the maximum v of BWVAC4(u,v) with u smaller than acU, and denote it as maxVL. Find the maximum v of BWVAC4(u,v) with u not less than acU, and denote it as maxVR.
8. The left half of BWVAC4(u,v) (with u smaller than acU) is recovered in two rounds if (maxVL-acV) is smaller than 50 mm. The first round is to compensate the morphological opening operation and the second round is to compensate the influence of the partial volume effect done similarly to the S landmark. The right half of BWVAC4(u,v) (with u not less than acU) is recovered in two rounds when (maxVR-acV) is smaller than 50 mm in a similar way.
9. When both (maxVL-acV) and (maxVR-acV) are smaller than 50 mm, the v coordinate of the I landmark is the biggest v coordinate of all object voxel in BWVAC4(u,v). If only one of (maxVL-acV) and (maxVR-acV) is smaller than 50, the v coordinate of the I landmark is the maximum v of object voxels from the side whose maximum object v coordinate is smaller than (50 mm+acV). When both of (maxVL-acV) and (maxVR-acV) are bigger than 50 mm, the v coordinate of the I landmark is the maximum v coordinate of all object voxels from the left or right side whose difference with acV is smaller.

Figs. 4a, b, and c show the binarized VAC (BWVAC1), processed foreground (BWVAC4), and the v coordinate of the I landmark overlaid on the original VAC slice.

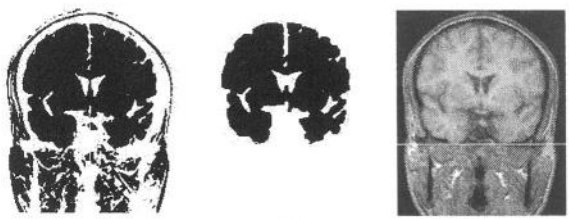


Fig. 4. Identification of the I landmark from processing the VAC plane: a) VAC being binarized into BWVAC1; b) BWVAC1 being processed into BWVAC4, and c) the horizontal line passing through the v coordinate of the extracted I landmark overlaid on the original VAC image

3 Results

The algorithm was implemented in C++ on Pentium 4 (2.6 GHz CPU). The 6 modified Talairach cortical landmarks were extracted within 2 seconds.

For each dataset, the 6 ground-truth landmarks were identified by a neuroanatomy expert (WLN) using the image editor Adobe Phtoshop on the dataset's AP, VAC, and VPC planes all with 1 mm^3 cubic voxels.

The range, average, and standard deviation of the landmark location errors for the A, P, L, R, I and S landmarks of the 62 datasets are listed in Table 1.

Table 1. The statistics of location errors for all the landmarks of the 62 datasets

	A	P	L	R	I	S
Range (mm)	0~2	0~3	0~2	0~1	0~2	0~3
Average (mm)	0.35	0.44	0.24	0.44	0.66	0.87
Standard deviation (mm)	0.58	0.62	0.47	0.50	0.63	0.76

The distribution of errors of all the 372 (62x6) landmarks is summarized in Table 2.

Table 2. The distribution of errors of all the landmarks

	0	1 mm	2 mm	3 mm
Number of landmarks	210	141	18	3
Percentage	56.5	37.9	4.8	0.8

4 Discussion

Our algorithm provides a way to automatically extract the modified Talairach cortical landmarks. The algorithm has been quantitatively validated against 62 MR datasets, including 18 brain web phantom T1-weighted datasets, 20 T1-weighted IBSR datasets, and 24 T1-weighted and SPGR clinical datasets.

The proposed algorithm is robust to noise and gray level inhomogeneity. For the T1-weighted phantom datasets with varying noise levels 0~9% and varying gray level inhomogeneity 0~40%, the landmark location error is mostly within 1 mm and only two landmarks have a location error of 2 mm. The algorithm can handle them well because the thresholds based on anatomical knowledge (in the form of range constraints) are robust to noise and gray level inhomogeneity as compared with other existing thresholding methods (Fig. 5).

The average location error is smaller than 1 mm and 94.2% of the extracted landmarks have a location error equal to 0 or 1 mm. The accuracy is achieved through compensating the connection between the cortex and the sagittal sinus/meninges, noise, gray level inhomogeneity, closeness of the cortex to optic nerves both spatially and in gray levels, and the partial volume effect.

The algorithm is fast, taking less than 2 seconds on Pentium 4 with 2.6 GHz CPU. This is mainly because only the three 2D images (the AP, VAC, VPC planes) are chosen to process instead of the whole 3D volumetric dataset. In addition, only simple operations like thresholding, seeding, simple morphological operations (erosion, dilation, opening and closing), distance transform are used. The algorithm can handle a wide spectrum of T1-weighted and SPGR clinical morphological volumes with

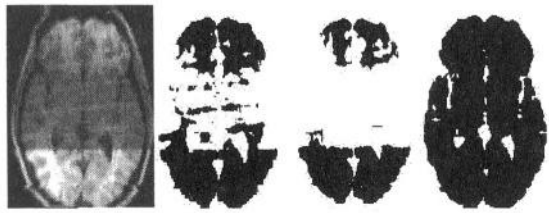


Fig. 5. Robustness to noise and inhomogeneity of the present method as compared with existing [2] and [8]: a) original AP plane of an IBSR dataset with serious inhomogeneity and noise; b) segmented AP plane based on [2]; c) segmented AP plane based on [8]; and d) segmented AP plane of the proposed algorithm.

various artifacts caused by a stereotactic frame as well as handles incomplete cortical surface.

1 Conclusion

We have proposed an algorithm to locate the modified Talairach cortical landmarks automatically, rapidly and robustly within 2 seconds on Pentium 4. The algorithm has been validated against 62 T1-weighted and SPGR MR datasets. The average landmark location error is below 1 mm: 0.35±0.58 mm for the A landmark, 0.44±0.62 mm for the P landmark, 0.24±0.47 mm for the L landmark, 0.44±0.50 mm for the R landmark, 0.66±0.63 mm for the I landmark, and 0.87±0.76 mm for the S landmark. The error distribution of all the landmarks is: 56.5% landmarks have no error, 37.9% landmarks have 1 mm error, 4.8% landmarks have 2 mm error, and 0.8% landmarks have 3 mm error.

Acknowledgement

The authors are grateful to the financial support by the Biomedical Research Council, Agency for Science, Technology and Research, Singapore.

References

1 Brummer ME, Mersereau RM, Eisner RL, Lewine RRJ, Automatic detection of brain contours in MRI data sets, IEEE Transactions on Medical Imaging (1993) 12(2): 153~66.
2 Cheng HD, Chen J, Li J, Threshold selection based on fuzzy c-partition entropy approach, Pattern Recognition (1998) 31: 857~70.
3 Cox RW, AFNI: software for analysis and visualization of functional magnetic resonance neuroimages, Computers and Biomedical Research (1996) 29: 162~73.
4 Hu QM, Two and three dimensional skeletonization, PCT Patent International Publication Number WO 02/058008 A1, 2002.

5 Hu QM, Hou ZJ, Nowinski WL, Methods and apparatus for binarizing images, PCT Patent Application BIL/Z/00461, 2003.
6 Hu QM, Nowinski WL, A rapid algorithm for robust and automatic extraction of the midsagittal plane of the human cerebrum from neuroimages based on local symmetry and outlier removal, Neuroimage (2003) 20(4): 2154~66.
7 Nowinski WL, Modified Talairach landmarks, Acta Neurochirurgica (2001) 143: 1045~57.
8 Otsu N, A threshold selection method from gray-level histograms, IEEE Transactions on Systems, Man and Cybernetics (1979) 9: 62~6.
9 Talairach J, Tournoux P, Co-planar stereotaxic atlas of the human brain. Thieme Medical Publisher, New York, 1988.
10 Verard L, Allain P, Travere JM, Baron JC, Bloyet D, Fully automatic identification of AC and PC landmarks on brain MRI using scene analysis, IEEE Transactions on Medical Imaging (1997) 16(5): 610~6.

Brain MR Image Segmentation Using Fuzzy Clustering with Spatial Constraints Based on Markov Random Field Theory

Yanqiu Feng, Wufan Chen

Laboratory of Medical Imaging, Department of Biomedical Engineering,
First Military Medical University, Guangzhou 510515, P.R. China.
foree@fimmu.com

Abstract. Unsupervised Fuzzy C-Means (FCM) clustering technique has been widely used in image segmentation. However, conventional FCM algorithm, being a histogram-based method when used in classification, has an intrinsic limitation: no spatial information is taken into account. This causes the FCM algorithm to work only on well-defined images with low level of noise. In this paper, a novel improvement to fuzzy clustering is described. The prior spatial constraint, which is defined as *refusable level* in this paper, is introduced into FCM algorithm through Markov random field theory and its equivalent Gibbs random field theory, in which the spatial information is encoded through mutual influences of neighboring sites. The algorithm is applied to the segmentation of synthetic image and brain magnetic resonance (MR) images (simulated and real) and the classification results show the new algorithm to be insensitive to noise.

1 Introduction

Unsupervised fuzzy clustering, especially fuzzy c-means algorithm (FCM), has been widely employed [1-7] in image segmentation. Based on minimum square error criterion, FCM algorithm can perform classification without need to estimate the density distribution, parametric or nonparametric, of the image. In addition, it is fairly robust and can be applied straightforwardly to multi-channel data. When used in image segmentation, however, FCM algorithm confronts a serious limitation: it does not incorporate any spatial information, which cause it to be sensitive to noise and imaging artifacts.

Mathematically, conventional FCM is formulated to minimize the following objective function, the sum of the errors between intensity at every pixel and the centroid of each class, with respect to the membership μ_{jk} and the centroid v_k

This work was supported by 973 Program of China (No: 2003CB716104) and the Grants from National Natural Science Fund of China (No.30130180)

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 188-195, 2004.

$$J_{FCM} = \sum_{j\in\Omega}\sum_{k=1}^{C} \mu_{jk}{}^{q} \left\| y_j - v_k \right\|^2 \quad (1)$$

Where y_j is observed intensity at pixel j, Ω is the set of pixel locations in the image, C is the number of clusters or classes. The membership functions are constrained to satisfy:

$$u_{jk} \in [0,1], \quad \sum_{k=1}^{C} u_{jk} = 1 \quad (2)$$

Here, the objective function (1) is minimized only when high values are assigned to pixels whose intensities are close to the centroid and low values are assigned to pixels whose intensities are distant from the centroid. The parameter q is the constant parameter that controls the degree of fuzziness of clustering result and satisfies $q>1$. The membership functions become increasingly fuzzy as q increases.

It can easily be seen from (1) that the objective function of FCM does not take into account any spatial information; i.e., segmentation is solely based on the histogram of image. This limitation will make the FCM algorithm exhibit sensitivity to noise in the observed image.

To overcome this limitation of FCM, one obvious way is to smooth the image before segmentation. However, standard smoothing filters can result in loss of important image details, especially at the transitory regions of image. More importantly, there is no way to rigorously control the trade-off between the smoothing and clustering. Another approach is to post-process the membership functions [1]. In [2], "multiresolution spatially constrained fuzzy membership function model" is applied to modify the prototype vectors. Spatial constraint is also enforced by incorporating "scale space vector" created by an area morphology operator [3].

Recently, approaches by directly modifying the objective functions have been proposed to increase the robustness of FCM to noise [4-6]. The distance is weighted in [4] by a factor based on the difference between the membership values of pixels in the neighborhood of the pixel. Penalty part is incorporated to the objective function (1) by D. L. Pham [5] to discourage unlikely or undesirable configurations according to the neighborhood of the pixels. In [6], a regularization term has been introduced to (1) to impose neighborhood effect. The above methods are claimed to be similar to the Markov random field theory, but they aren't directly based on MRF, the very efficient and competent theory to describe the spatial context information in image analysis. And the methods have to be confronted with the problem of selecting the parameter that controls the balance between conventional part and added part for spatial constraint. In [5], cross-validation has been utilized to select the parameter of regularization. Although subsampling of pixels has been implemented to construct the validation set, the computation load is still very high.

MRF modeling and its application in image segmentation have been investigated by many researchers [7-8]. It has been shown that MRF prior can improve the performance of segmentation.

Based on MRF theory and its equivalent GRF theory, we introduce the spatial context constraint into the objective function of FCM. Minimize the new objective func-

tion according to the zero gradient condition, thus a novel GFCM algorithm is proposed which can handle both the grayscale and spatial information while segmenting. The rest of this paper is organized as follows: Section 2 presents the mathematical details of MRF and GRF theory. Section 3 describes the GFCM algorithm. Experimental results of image segmentation and comparison with other methods are given in Section 4, followed by conclusions in the final section.

As to the problem of brain MR images segmentation, we focus on segmenting normal brain image, with non-brain parts of the images having been removed in advance, into three kinds of tissues: Gray matter (GM), White Matter (WM), and CSF.

2 MRF Theory and GRF Model

Let $X=\{x=(x_1,\cdots,x_N) \mid x_i \in L, i \in S\}$ be a family of random variables defined on the set of sites $S=\{1,2,\cdots N\}$, in which each random variable X_i takes a value in the set of labels $L=\{1,2,\cdots,l\}$. The family X is called a random field [7].

2.1 MRF Theory

MRF theory provides a convenient and consistent way to model context dependent constraint through neighborhood system: $N=\{N_i, i \in S\}$ [7], where N_i is the set of sites neighboring i, and has the properties: (1) $i \notin N_i$; (2) $i \in N_j \Leftrightarrow j \in N_i$.

X is said to be a Markov random field (MRF) on S with respect to a neighborhood system N if and only if the following two conditions are satisfied:

$$P(x) > 0, \forall x \in X \tag{3}$$

$$P(x_i \mid x_{S-\{i\}}) = P(x_i \mid x_{N_i}) \tag{4}$$

The Markovianity depicts the local characteristics of X: a label interacts with only its neighboring labels. In other words, only neighboring labels have direct interactions with each other. According to the Hammersley-Clifford theorem [7], an MRF can be equivalently characterized by a Gibbs distribution.

2.2 GRF Theory

A set of random variables X is said to be a *Gibbs random field* (GRF) on S with respect to N if and only if its configurations obey a *Gibbs distribution.* A Gibbs distribution takes the following form [7]:

$$P(x) = \exp(-U(x)) / Z \tag{5}$$

where Z is a normalizing constant called the *partition function* and $U(x)$ is the *energy function.* The energy is the sum of *clique potentials* $Vc(x)$ over all possible cliques:

$$U(x) = \sum_{c \in C} V_c(x) \tag{6}$$

A clique is defined as a subset of sites in *S* in which every pair of distinct sites are neighbors. The value of *Vc(x)* depends on the local configuration on the clique *c*.

2.3 Multi-Level Logistic Model (MLL)

By choosing different clique potential function $V_c(x)$, a wide variety of distributions can be formulated as Gibbs distributions. In MLL models [7][9], the potentials for cliques containing more than one site are defined as:

$$V_c(x) = \begin{cases} -\beta_c & \text{if all sites in c have the same label} \\ \beta_c & \text{otherwise} \end{cases} \tag{7}$$

$\beta_c > 0$ is a constant depending on the type of clique. This encourages neighboring sites to have the same class label. Homogeneity is imposed on the model by assigning the same potential function to all cliques of a certain type, independent of their positions in the image.

In this paper, we use a special MLL model [8] that considers only two-site cliques so that the energy function can be written as:

$$V_2(x_i - x_j) = \beta[1 - \delta(x_i - x_j)] \tag{8}$$

$\delta(\cdot)$ is the Kronecker delta fuction and β is the penalty against non-equal labels on two-site cliques and inversely proportional to the signal noise ratio (SNR) of MRI data.

3 GFCM Algorithm

Given the basic theory on GRF model in last section, the definition of *refusable level* is proposed below.

Propositions 1: Given one pixel *j* and the set of its neighboring sites *Nj,* the prior probability $P_j(k)$ of labeling the pixel to $k \in L$ can be calculated in term of the Gibbs model presented in Section 2, $1 - P_j(k)$ can be considered as the resistance of neighbors *Nj* to assigning pixel *j* the label *k.* The resistance is defined as *refusable level* in this paper.

Because *refusable level* take a value between the 0 and 1 and represent the spatial constraints, it can be introduced into the objective function (1) of FCM as following,

$$J_{GFCM} = \sum_j \sum_{k=1}^{C} \mu_{jk}^q \left(1 - P_j(k)\right) \left\| y_j - v_k \right\|^2 \tag{9}$$

where $1-P_j(k)$ is calculated from the hard maximum membership segmentation during the iteration.

In order to minimize the objective function (9), membership values are assigned to the pixel not only according to the distance from the centroid, but also taking into account the resistance of the neighboring pixels to the label. When minimizing (9), the mutual influences and collective role between *refusable level* and intensity distance are described in detail as follows:

At pixel j, when the *refusable level* $1-P_j(k)$ of the neighboring pixels Nj to label k is low, the high resistance caused by large distance of the intensity to centroid can be tolerated by the low *refusable level.* As a result, a high membership may be assigned to pixel j. If *refusable level* $1-P_j(k)=0$, it will tolerate all resistance caused by intensity no matter how distant the intensity is from the centroid and the pixel will definitely be assigned label k. While the reusable level $1-P_j(k)$ of the neighboring pixels to the label k is high, it will give the intensity distance a large weight in the objective function and will encourage assigning a low value to label k.

Using Lagrange multipliers to impose the constraint in (2) and evaluating the centroids and membership functions that satisfy a zero gradient condition yield the two necessary conditions for J_{GFCM} to be at a minimum.

$$\mu_{jk} = \frac{\left(\left\|y_j - v_k\right\|^2 \left(1 - P_j(k)\right)\right)^{-1/(1-q)}}{\sum_{l=1}^{C} \left(\left\|y_j - v_k\right\|^2 \left(1 - P_j(k)\right)\right)^{-1/(1-q)}} \tag{10}$$

$$v_k = \frac{\sum_{j \in \Omega} \mu_{jk}^{\,q} \left(1 - P_j(k)\right) y_j}{\sum_{i,j} \mu_{jk}^{\,q} \left(1 - P_j(k)\right)} \tag{11}$$

The discrete steps of Gibbs Fuzzy C-means (GFCM) Algorithm are as follows:

1. Initial estimates of the centroids and initial segmentation;
2. Calculate the prior probability *Pj*(k) in term of (5);
3. Update the membership functions according to (10);
4. Calculate centroids using (11);
5. Go to step (2) and repeat until convergence.

4 Experiments and Discussions

In this section, the segmentation results of GFCM algorithm are shown. In these experiments, we set *C=4* and *q=2*. The algorithms were run on a 1.2G PC system. For an image with size of 256*256, execution time of GFCM algorithm is about 5s and standard FCM algorithm requires about 3s.

The results of applying GFCM algorithm to a synthetic test image is shown in Figure 1. The test image contains intensity values 0, 85, 170 and 255, and the image size is 256*256 pixels. White Gaussian noise with a standard deviation of 30 was added to the image. Figure 1a shows the test image. Figure 1b and 1c show the results of

maximum membership classification produced by standard FCM and GFCM, respectively. It can be seen that the result of GFCM classification is less speckled and smoother except that there is some faint distortion at edges. Therefore, the GFCM classification, under the spatial smoothing constraints of the neighborhood system, is much more robust than the traditional FCM classification.

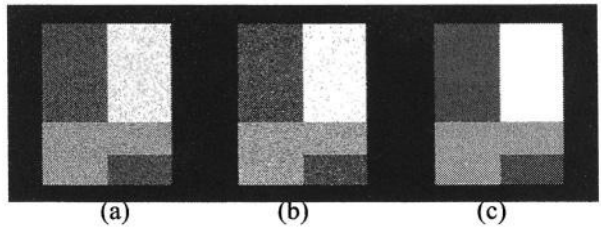


Fig. 1. Comparison of segmentation results on a two-dimensional test image: (a) the test image, (b) FCM classification, (d) GFCM classification.

Figure 2 shows the application of GFCM to a simulated MR image from the Brainweb database. Figure 2a shows the original simulated image, 9% noise and no inhomogeneity. Figure 2b and 2c shows the segmentation result of standard FCM and GFCM, respectively and the ground truth of segmentation is shown in figure 2d. Obviously Figure 2c is much closer to the ground truth than figure 2b and the smoother appearance of the GFCM result is evident.

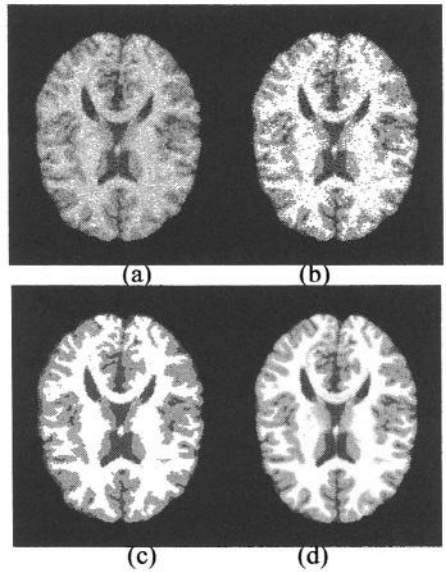


Fig. 2. Comparison of segmentation results on a simulated MR image: (a) original image, (b) using FCM, (c) using GFCM, (d) the true segmentation used to simulate the MR image

The correct classification rates (CCR) of applying several different algorithms to the simulated MR images for different levels of noise were shown in Table 1. MFCM

is the modified FCM algorithm [6] and PFCM is the penalized FCM algorithm [5]. With the increase of noise level, the segmentation result of standard FCM degrades rapidly. While the fuzzy clustering algorithms with spatial constraint such as GFCM, PFCM and MFCM, can overcome the problem caused by noise. Overall, the three kinds of improved FCM algorithm produce comparable results, i.e. our GFCM algorithm provide another novel approach to improve the performance of conventional FCM algorithm.

Table 1. Correct Classification Rates of Different Methods Applied on Simulated MR Data

	Noise level (%)			
	3%	5%	7%	9%
FCM	92.12	88.94	84.25	79.23
GFCM	96.86	95.67	95.34	95.12
MFCM	97.18	96.63	95.82	94.33
PFCM	94.27	93.52	93.13	92.63

Figure 3 shows the application of RFCM to real MR images taken from IBSR. The algorithm has incorporated the bias field correction as described in [6]. Novel methods is also addressed in [10][11][12] to correction for the intensity inhomogeneity in MR images. Figure 3a shows the original T1 weighted brain MR images, Figure 3b shows the estimated bias field and Figure 3c shows the Segmentation result of GFCM with bias field correction. The manual segmentation by medical expert is shown in figure 3d. Obviously, GFCM algorithm with the bias field correction produces classification comparable to manual results of expert.

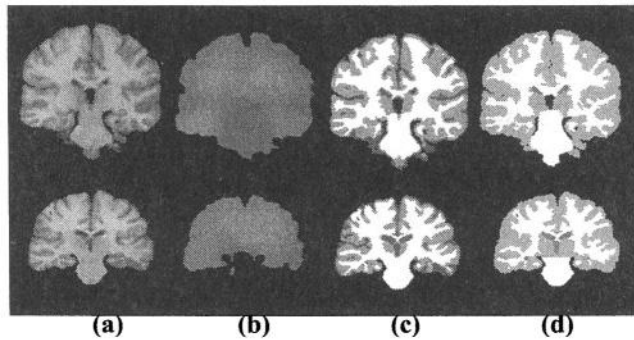


Fig. 3. GFCM applied to real T1-weighted MR images from IBSR: (a) original images, (b) estimated bias fields, (c) segmentation results of GFCM, (d) manual segmentation results

5 Conclusion

In this paper, we have described a novel extension of FCM, based on the Markov Random Fields theory, to incorporate spatial constraints. The spatial information in GFCM is extracted from the label set during segmentation. Comparison is also pre-

sented between our GFCM algorithm and traditional FCM, penalized FCM and Modified FCM on synthetic image and simulated MR images. The GFCM produces comparable results as PFCM and MFCM, while GFCM is of complete mathematical background with more promising future improvement

References

1. Y. A. Tolias and S. M. Panas: On applying spatial constraints in fuzzy image clustering using a fuzzy rule-basedsystem. IEEE Signal Process. Lett. 5(10), 1998, 245–247
2. Y. A. Tolias and S. M. Panas: Image segmentation by a fuzzy clustering algorithm using adaptive spatially constrained membership functions. IEEE Trans. Systems, Man, Cybernet. A 28(3), 1998, 359–369
3. S. T. Acton and D. P. Mukherjee: Scale space classification using area morphology. IEEE Trans. Image Process, 9(4), 2000, 623–635
4. A. W. C. Liew, S. H. Leung, and W. H. Lau: Fuzzy image clustering incorporating spatial continuity. IEE Proc. Visual Image Signal Process, 147(2), 2000, 185–192
5. D. L. Pham: Spatial Models for Fuzzy Clustering. Computer Vision and Image Understanding 84, pp. 285-297, 2001
6. M. N. Ahmed, Sameh M. Yamany, Nevin Mohamed, Aly A. Farag and Thomas Moriarity: A Modified Fuzzy C-Means Algorithm for Bias Field Estimation and Segementation of MRI Data. IEEE trans. On Medical Imaging, 21(3): 193-199, 2002
7. Stan Z. Li: Markov Random Field Modeling in Image Analysis. Springer, 2001, ISBN 4-431-70309-8
9. Y. Zhang, M. Brady, and S. Smith: Segmentation of Brain MR Images through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm. IEEE Trans. Medical Imaging, 20(1):45-57, 2001
10. S. German and D. German: Stochastic relaxation, Gibbs distribution, and the Bayesian Restoration of Images. IEEE Trans. Patter Anal. Machine Intell., Vol. PAMI-6, No. 6, pp. 721-741, 1984
11. Dzung L. Pham, Jerry L. Prince: Adaptive fuzzy segmentation of magnetic resonance images. IEEE Trans. On Medical Imaging, 18(9): 737-752[2], 1999
12. C. Z. Zhu, T. Z. Jiang: MultiContext fuzzy clustering for separation of brain tissues in magnetic resonance images. NeuroImage, Vol.18, No.3, pp. 685-696, 2003

Anatomy Dependent Multi-context Fuzzy Clustering for Separation of Brain Tissues in MR Images

C.Z. Zhu, F.C. Lin, L.T. Zhu and T.Z. Jiang

National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, P.R. China
{czzhu, fclin, ltzhu, jiangtz}@nlpr.ia.ac.cn

Abstract. In a previous work, we proposed multi-context fuzzy clustering (MCFC) method on the basis of a local tissue distribution model to classify 3D T1-weighted MR images into tissues of white matter, gray matter, and cerebral spinal fluid in the condition of intensity inhomogeneity. This paper is a complementary and improved version of MCFC. Firstly, quantitative analyses are presented to validate the soundness of basic assumptions of MCFC. Carefully studies on the segmentation results of MCFC disclose a fact that misclassification rate in a context of MCFC is spatially dependent on the anatomical position of the context in the brain; moreover, misclassifications concentrate in regions of brain stem and cerebellum. Such unique distribution pattern of misclassification inspires us to choose different size for the contexts at such regions. This anatomy-dependent MCFC (adMCFC) is tested on 3 simulated and 10 clinical T1-weighted images sets. Our results suggest that adMCFC outperforms MCFC as well as other related methods.

1 Introduction

In general, white matter (WM), gray matter (GM) and cerebral spinal fluid (CSF), are three basic tissues in the brain. Brain tissue segmentation of MR images means to identify the tissue type for each point in data set on the basis of information available from both MR images and neuroanatomical knowledge. It's an important processing step in many medical research and clinical applications, such as quantification of GM reduction in neurological and mental diseases, cortex segmentation and analysis, surgical planning, multi-modality images fusion, functional brain mapping [2,9,10]. Unfortunately, intensity inhomogeneity, caused by both MR imaging device imperfections and biophysical properties variations in each tissue class, result in various MR signal intensities for the same tissue class at different locations in the brain. Hence intensity inhomogeneity is a major obstacle to any intensity based automatic segmentation methods and has been investigated extensively [1,4,5,6,7,8,11]. To address this issue, multi-context fuzzy clustering (MCFC) method had been proposed [11] on the basis of a local tissue distribution model in our previous work. In this paper, anatomy-dependent MCFC (adMCFC) is proposed to refine and improve the original MCFC.

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 196-203, 2004.

The local tissue distribution model and MCFC method are briefly summarized in Section 2. Section 3 presents adMCFC followed by experimental results in Section 4. The final section is devoted to discussion and conclusions.

2 Original MCFC Method

Clustering context is a key concept for the local tissue distribution model and MCFC. A context is a subset of 3-D MRI brain volume and the size of a context is defined as the number of pixels in the context. Highly convoluted spatial distributions of the three different tissues in the brain inspired us to propose the local tissue distribution (LTD) model. Given a proper context size, LTD model in any context consists of the following three basic assumptions:

(1) Each of the three classes of tissues exists with considerable proportion.

(2) All pixels belonging to same tissue class will take on similar ideal signal intensities.

(3) Bias field is approximately a constant filed.

As a whole, it is conflictive to choose context size for the three assumptions simultaneously. The following quantitative analyses are presented as a complementary explanation to validate the soundness of the assumptions. The simulated T1-weighted data as well as the corresponding labeled brain from the McConnell Brain Imaging Center at the Montreal Neurological Institute, McGill University [3] were used through the paper.

Firstly, an index of fractional anisotropy (FA) was presented to describe difference in proportions of the three tissue classes in a given context.

$$FA = \frac{\sqrt{3}\cdot\sqrt{(n_w-\overline{n})^2+(n_g-\overline{n})^2+(n_s-\overline{n})^2}}{\sqrt{2}\cdot\sqrt{n_w^2+n_g^2+n_s^2}} \tag{1}$$

Where n_w, n_g and n_s are the number of pixels in a context belonging to WM, GM, and CSF respectively and $\overline{n}$ is the average. Given normalized context size (NCS), N contexts or brain regions were uniformly sampled in the labeled brain image and the averaged FA among N contexts was defined as a measure of proportion difference for the given NCS [11]. When pixel amount were same for all the three tissues in a context, FA reached zero as the minimum; when the amount differences among the three tissues became more significant FA would increased. As a function of NCS, FA was plotted in Fig.1 where FA is decreasing when the context became larger. Accordingly, larger context size could guarantee assumption (1) better.

As for assumption (2) the centers of ideal signal distributions (CISD) were calculated for both WM and GM in each sampled context with given NCS in the 3-D simulated T1-weighted data without any noise and bias field imposed. Profiles of CISD of GM in contexts at different positions in the brain were plotted in Fig 3 (a) ~ (d) corresponding to NCS of 0.02,0.06, 0.10 and 0.18 respectively.

In any one of the 4 figures, CISD are various in different brain regions, which implies the intrinsic variations of biophysical properties in GM. Moreover, such variation of CISD gradually decreases when the NCS becomes larger and larger, which

suggests local differences in biophysical properties of different GM structures are gradually vanished. Accordingly assumption (2) required the context as small as possible to keep the truth. As for assumption (3), however, the smaller, the better.

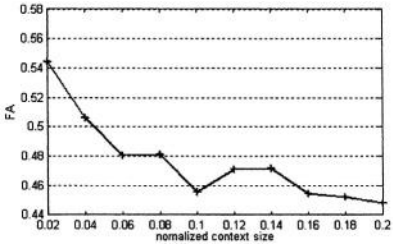


Fig. 1 FA distribution with NCS

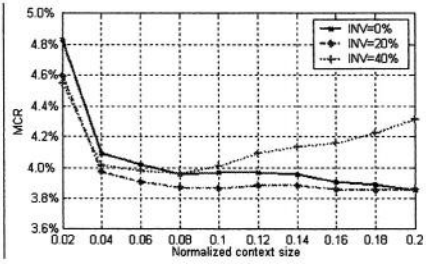


Fig. 2 MCR distribution with NCS

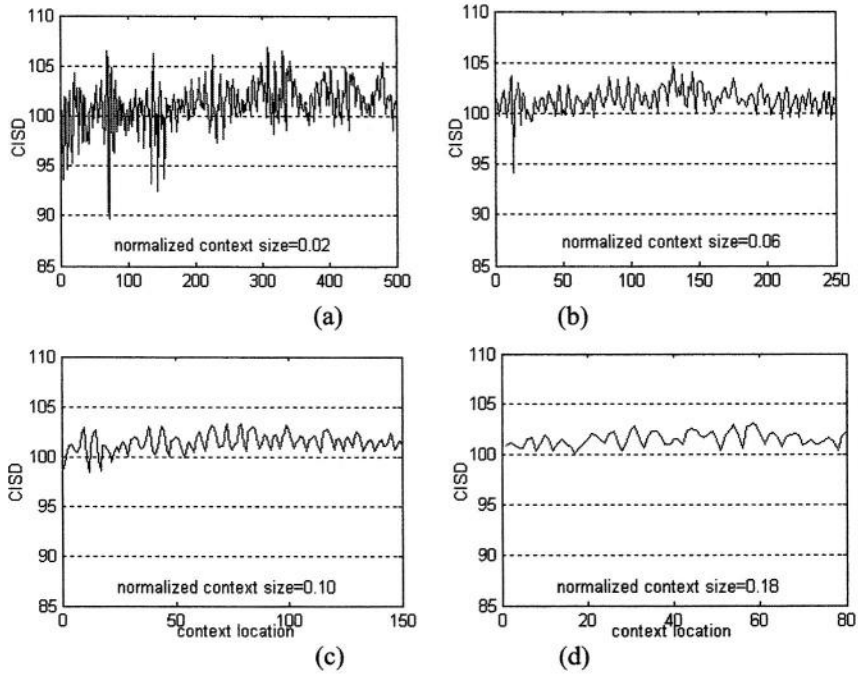


Fig. 3. GM CISD distributions with NCS= 0.02(a); 0.06 (b); 0.10 (c) and 0.18 (d)

Therefore it is conflictive to choose a proper context size since assumption (1) asks the size as larger as possible, while the other two assumptions require the opposition. As a function of NCS, misclassification rates (MCR) of MCFC on simulated data with 3% noise were plotted in Fig.2.in case of 0%, 20% and 40% bias field (INV) respectively. We can see that 0.06, a tradeoff between the two conflictive requirements, yielded satisfying results.

Given NCS, MCFC includes two stages: multi-context fuzzy clustering and information fusion. Firstly, multiple clustering contexts are generated for each voxel and fuzzy clustering is independently carried out in each context to calculate the membership of the voxel to each tissue class. The memberships can be regarded as soft decisions made upon the information from each information source, say the context. Then in stage 2, all the soft decisions are integrated as the final results. Implementation details of MCFC can be found in [11].

3 Anatomy-Dependent MCFC

Carefully studies on the MCR of each context resulted in an interesting finding that MCR in contexts varied at different position in the brain and most of the errors concentrated in the area of brain stem and cerebella as shown in Fig. 4. The finding seemed similar for all the three data sets (3% noise and 0%, 20% and 40% bias fields).

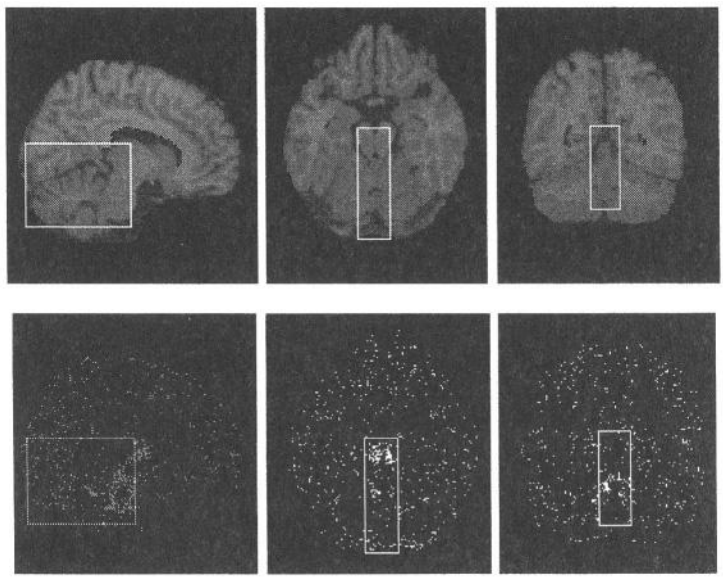


Fig.4 Original simulated T1-weighted MR data and the misclassified pixels. Box indicates the area with concentrated misclassification.

Quantitatively analysis in FA suggested that higher FA made assumption (1) not well guaranteed, which was, at least, part of the reason to concentrated misclassification in such area. We enlarged contexts in these regions by an enlarging coefficient. As a function of enlarging coefficient, averaged FA among the enlarged contexts were calculated and plotted in Fig. 5. We can see that FA in such area does decrease with the enlarging coefficient so that assumption (1) could be more correct. The implementation of adMCFC can be summarized as follows:

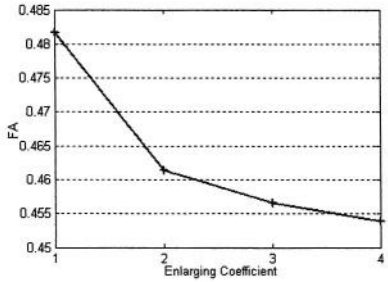


Fig. 5. FA and enlarging coefficient

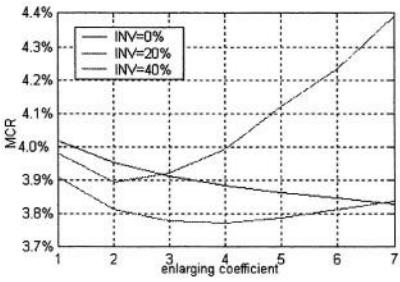


Fig. 6 MCR and enlarging coefficient

Step 1 Find the enlarging anatomic area in target brain images
In practice, a binary mask could be rough created either by a manually drawing covering brain stem and cerebella or by a rigid registration from the template as shown in Fig. 4 to the target.

Step 2 Perform modified MCFC with a given NCS.
During context window with NCS moving through the target image, context center is tested whether in the enlarging mask or not. If yes, enlarge the context by multiplying the NCS with a given enlarging coefficient; If not, keep original NCS unchanged. Do step 2 until all the contexts were processed.

In adMCFC, a proper enlarging coefficient is very important. Given NCS = 0.06 as in [11], three MCR curves were calculated and plotted as the function of the enlarging coefficient in Fig. 6 on the three sets of simulated T1-weighted MRI data respectively. When enlarging coefficient was set between 1 and 4, the MCR became smaller for all the intensity inhomogenieties conditions and the best classification results occurred at slightly different enlarging coefficients for the three conditions. In this work, we chose 3 as the enlarging coefficients in all experiments.

4 Experiments

4.1 Evaluation with 3-D Simulated MRI Data

adMCFC as well as MCFC and other related methods were tested on the same simulated MR data as in [11] and the results were listed in Table 1. Both adMCFC and MCFC were significantly robust to increased bias field than FCM and FM-AFCM in [6]. Moreover, MCR of adMCFC was lower than that of MCFC in each level of bias field. Additionally, adMCFC outperformed MCFC in the masked area can visually demonstrated in Fig.7 where the misclassification pixels were obviously reduced.

Table 1. MCR from simulated data

Method	INV=0%	INV=20%	INV=40%
FCM	4.020%	5.440%	9.000%
FM-AFCM	4.168%	4.322%	4.938%
MCFC	4.020%	3.909%	3.979%
adMCFC	3.915%	3.780%	3.922%

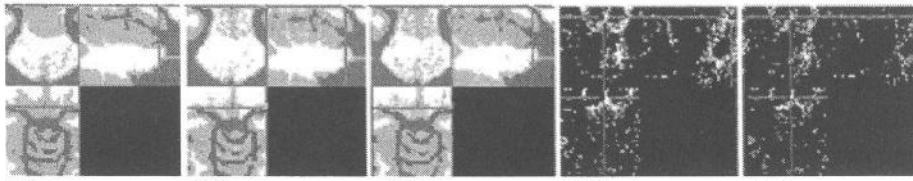


Fig. 7. Segmentation results. (a) True model (b) MCFC result (c) adMCFC result (d) MCFC misclassification (e) adMCFC misclassification

4.2 Evaluation with Real T1-weighted MRI Data

T1-weighted MRI data of 10 normal subjects (SIEMENS 1.5T, image size: 256×256×128, resolution: 1.17mm × 1.17mm× 1.25mm) were used to validate adMCFC. Fig. 8 showed the results from both adMCFC and FCM on the MR image of one of the 10 subjects. The bias field and corresponding misclassifications can be easily detected at the top part in the original MR image and segmentation results of FCM. But adMCFC yielded a much better result. Such improvement can also be demonstrated with 3D rendering of the segmented WM in Fig 9 where WM loss is very significant at the top area in the results of FCM.

5 Discussions and Conclusion

In this work we have qualitatively described the requirements of our LTD model and presented a improved MCFC method to separated brain tissue in T1-weighted MR images with more accuracy than MCFC as well as other related methods in the condition of bias field and biophysical properties variations. It is difficult for a fixed context size to guarantee the assumptions of LTD in all contexts because of the complexity of the brain. While adMCFC can determine the size of a context according to its anatomic position to result in a lower misclassification rate than original MCFC can do.

There are several issues to study further to improve the performance of adMCFC, such as the relationship between context size and enlarging coefficient, the shape and size of the mask to enlarge context. Please note that, we corrected a minor mistake in the MCFC algorithm so that we obtained slightly different results in Table 1 and Fig. 2 from those in [11].

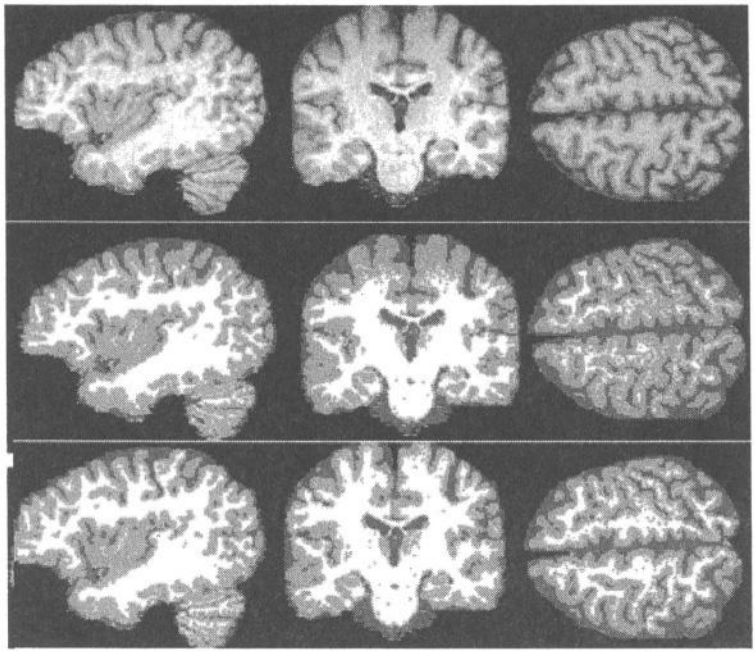


Fig. 8. Segmentation results. Original images (first row); FCM results (second row) and adMCFC results (third row)

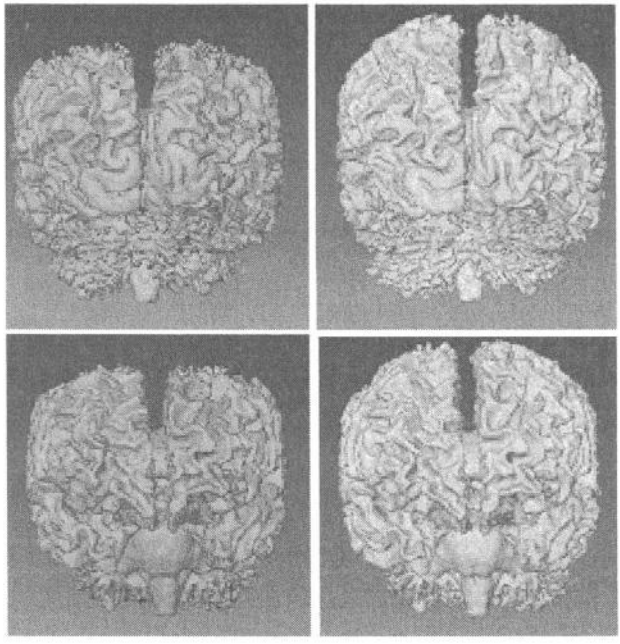


Fig. 9. 3D rendering of segmented WM from two view angles (top and bottom row). FCM results (left column) and adMCFC results (right column)

References

1. Arnold, J. B., Liow, J.-S., Schaper, K. A., et al. Qualitative and Quantitative Evaluation of Six Algorithms for Correcting Intensity Nonuniformity Effects. NeuroImage (2001) 13: 931–934.
2. Clarke, L. P., Velthuizen, R. P., Camacho, M. A., Heini, J. J., Vaidyanathan, M, Hall, L. O., Thatcher, R. W. and Silbiger, M. L. MRI Segmentation: Methods and Applications. Mag. Res. Imag. (1995) 13: 343-368.
3. Collins, D. L., Zijdenbos, A. P., Kollokian, V., Sled, J. G.., Kabani, N.J., Holmes,C.J., Evans, A.C. Design and Construction of a Realistic Digital Brain Phantom. IEEE Trans. Med. Imag. (1998) 17: 463–468.
4. Guillemaud, R. and Brady, M. Estimating the Bias Field of MR Images. IEEE Trans. Med. Imag. (1997). 16: 238-251.
5. Likar B., Viergever, M. A. and Pernus F., Retrospective Correction of MR Intensity Inhomogeneity by Information Minimization, IEEE Trans. Med. Imag. (2001). 20: 1398-1410.
6. Pham, D. L. and Prince J. L. Adaptive Fuzzy Segmentation of Magnetic Resonance Images. IEEE Trans. Med. Imag. (1999) 18: 737-752.
7. Shattuck, D. W., Sandor-Leahy, S. R., Schaper, K. A., Rottenberg, D. A. and Leahy, R. M. Magnetic Resonance Image Tissue Classification Using a Partial Volume Model. NeuroImage. (2001) 13: 856-876.
8. Sled J.G., Zijdenbos A.P., and EvansA.C., A Nonparametric Method for Automatic Correction of Intensity Nonuniformity in MRI Data, IEEE Trans. Med. Imag. (1998). 17: 87-97.
9. Suri, J. S., Setarehdan, S. K. and Singh, S. Advanced Algorithmic Approaches to Medical Image Segmentation: state-of-the-art applications in cardiology, neurology, mammography and pathology. Spring-Verlag London Berlin Heidelberg, (2002).
10. Worth, A. J., Markis, N., Caviness, V. S. and Kennedy, D. N. et al. Neuroanatomical segmentation in MRI: Technological Objectives. International J. of Pattern Recognition and Artificial Intelligent (1997) 11:1161-1187.
11. Zhu C.Z.and Jiang T. Z., “Multicontext fuzzy clustering for separation of brain tissues in magnetic resonance images”, NeuroImage, Volume 18, Issue 3, (2003) Pages 685-696.

Visual Search in Alzheimer's Disease — fMRI Study

Jing Hao[1], Kun-cheng Li[1], Ke Li[2], De-xuan Zhang[3], Wei Wang[1], Bin Yan[2], Yan-hui Yang[1], Yan Wang[3], Qi Chen[3], Bao-ci Shan[2], and Xiao-lin Zhou[3]

[1] Department of Radiology, Xuanwu Hospital, Capital University of Medical Sciences, Beijing 100053, China
[2] Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100039, China
[3] Department of Psychology, Peking University, Beijing 100034, China

Abstract. The aim was to investigate the neural basis of visual attention deficits in Alzheimer's disease (AD) patients using functional MRI. Thirteen AD patients and 13 age-matched controls participated in the experiment of two visual search tasks, one was a pop-out task, the other was a conjunction task. The fMRI data were collected on a 1.5T MRI system and analyzed by SPM99. Both groups revealed almost the same networks engaged in both tasks, including the superior parietal lobule (SPL), frontal and occipito-temporal cortical regions (OTC), primary visual cortex and some subcortical structures. AD patients have a particular impairment in the conjunction task. The most pronounced differences were more activity in the SPL in controls and more activity in the OTC in AD patients. These results imply that the mechanisms controlling spatial shifts of attention are impaired in AD patients.

1 Introduction

Alzheimer's disease (AD) was considered as a dementia characterized by global cognitive impairment. Amnesia has long been recognized as a primary manifestation and is the core symptom for the clinical diagnosis of probable AD [1]. However, there has been a suggestion that attention is impaired early in the course of AD [2]. Until recently, there has been a relative paucity of experimental studies about attentional functions in AD. Attentional impairments have been revealed in many studies of attentional capacities including both auditory [3] and visual [4] selective processing, visual search [5] and attention shifting [6]. However, other studies have showed no marked deficits in detecting, shifting to and engaging target items [7,8]. There is therefore a continuing debate concerning the status of attentional functions in AD patients. In the present study, we used computer-presented visual search task to examine whether an attentional deficit exists in AD, We also intended to investigate the neural basis of visual attention deficits with functional magnetic resonance imaging (fMRI).

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 204–212, 2004.

2 Method

2.1 Subjects

Thirteen patients (mean age, 62.6 ± 7.8; 8 females) suffering from mild to moderate AD were recruited from our outpatient memory disorder unit (diagnosed according to NINCDS-ADRDA [1] and ICD-10 criteria [9]). Relevant medical disorders (asides from AD) were excluded. The severity of cognitive impairment was assessed using the Mini Mental State Examination (MMSE) [10] (group mean score, 14.3 ± 8.2). The control group consisted of 13 healthy subjects, matched to the patient group in age (mean age 64.5 ± 6.7) and gender (7 females). All controls had a score 27.8 ± 2.6 points in the MMSE and no pathological changes in T1 and T2 structural cranial MR images. They had no psychiatric, neurological, or cardiovascular disease history and did not use psychotropic drugs. All subjects provided written informed consent prior to participation.

2.2 Experimental Paradigm

The visual search tasks were generated on a personal computer using the presentation software package. Subjects were stabilized against bulk head movements using custom-made foam pads. They viewed the stimuli through a mirror mounted on the head coil. All participants performed two tasks. One was a pop-out single feature task, detecting a vertical target bar among horizontal distractors, regardless of their color (Fig.1a). The other was a conjunction task, in which the target is defined by a conjunction of features (color and orientation) and the performance depends on some shifting of attention (Fig.1b). The visual display subtended a maximum size of 12° horizontally and 8° degrees vertically. In both visual search conditions, subjects were instructed to respond to the presence or absence of the target as quickly as possible with right-hand button press, while avoiding errors. Three stimulus set size (4, 8 and 12) were randomly varied from trial to trial and the target was present in 50% of trials. At the beginning of each trial a central fixation cross(+) was presented for 500 ms, followed by the array of visual stimuli for 3000 ms. A blank interval of 1000 ms intervened between trials.The functional scan followed a classic block design where the stimuli were presented in six blocks (54 s of each), alternating with fixation periods of 27 s. Reaction time (RT) and correctness of response were recorded. To ensure visual search performance under steady fixation, subjects underwent about 1-hr training sessions.

2.3 MR Imaging

The fMRI examination was performed using a 1.5-T MRI system (Siemens Sonata, Germany). For functional imaging, 16 slices [(slice thickness=5 mm, slice gap=1 mm; flip angle (FA)=90°; matrix size=64 × 64; field of view(FOV)=220 mm×220 mm), were acquired using a gradient-echo echo-planar imaging (GE-EPI) sequence with a repetition time (TR) of 4500 ms, and an echo time (TE) of

50 ms. Each functional time series consisted of 108 volumes and lasted 486 s. Additionally, structural three-dimensional data sets were acquired in the same session using a T1-weighted sagittal MP-RAGE sequence (TR=1900 ms, TE=3.93 ms; matrix=448 × 512; thickness=1.70 mm, gap=0.85 mm; FOV=250 mm×250 mm).

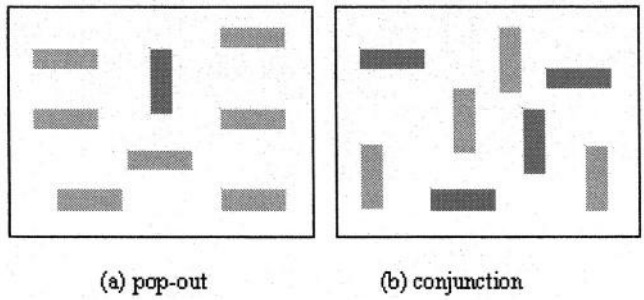


Fig. 1. Diagrammatic representation of the two visual search tasks used in this study (dark and light bars designate red and green bars, respectively)

2.4 Data Analysis

Reaction Time (RT) and correctness of response were compared across conditions (pop-out and conjunction) and between groups (controls and AD patients) using two-way ANOVAs. For each of the four experimental conditions (controls or AD patients doing pop-out or conjunction task), a two-way analysis of ANOVA was used with distractors and target presence or absence as main factors. Additionally, RT×distractor set size slope was calculated.

SPM99 was used for imaging data preprocessing and statistical analysis [11,12]. The statistical effects of task conditions and subjects were estimated according to the general linear model applied to each voxel in brain space. Statistical comparisions between experimental factors were based on the fixed-effects model. The different activations between groups and within group were analyzed using 2-way ANVOA. The statistical threshold was set at $P < 0.001$ uncorrected and a cluster threshold of greater than 10 voxels.

3 Results

3.1 Behavioral Data

Both groups performed well and relatively few errors were made. The results showed that search rates in the pop-out condition were similar in both the two groups, However, AD patients searched significantly more slowly compared to the controls in the conjunction condition($P < 0.05$).

We obtained intercept (in milliseconds) and slope (in milliseconds per distractor) values from linear regression analyses with median correct RT set against the array size(4, 8 and 12) for groups (AD and controls) on both tasks. These values are shown in Table 1. In the conjunction condition, for normal controls, there was a significant interaction (2-way ANOVA, $P < 0.05$) between set size and the presence or absence of the target, i.e. the number of distractors had a different effect on present and absent responses. In contrast, for AD patients, there was no such interaction (2-way ANOVA, $P > 0.05$). Normal controls showed the classical 2 : 1 target present/absent response pattern (target present slope: 20ms/item; target absent slope: 44ms/item); Patients with AD, however, showed a different pattern, in that the difference between target present and target absent slope was much smaller (target present slope, 35ms/item; target absent slope, 41 ms/item).

Table 1. Intercept, Slope and r^2 Values for the Two Groups on the Popout and Conjunction Tasks (Note. TP: target present; TA: target absent)

	Simple feature search task		Conjunction feature search task	
	controls	AD	controls	AD
TP-slope	3.25	0.75	20.50	35.25
TP-intercept	804.00	772.33	750.33	775.33
TP-r^2	0.99	0.30	0.92	0.98
TA-slope	-1.75	2.50	44.50	41.75
TA-intercept	739.00	720.00	828.00	764.00
TA-r^2	0.20	0.04	0.99	0.99

3.2 Imaging Data

Group average activations from the search tasks were shown in Fig.2. Mean locations in Talairach Space and volumes were given in Table 2 and 3. The analysis of group for AD and controls revealed that the cerebral networks involved in both tasks were almost the same, including the parietal lobule (SPL), frontal and occipito-temporal cortical regions (OTC), primary visual cortex and some subcortical structures.

There were remarkable differences in the extent of activation of these brain regions between patients and controls (Fig. 2). The less activation in AD patients was demonstrated in the bilateral parietal lobes and right frontal regions, while additional activation was found in left frontal lobes and right occipito-temporal cortical regions in the conjunction condition. There was a double dissociation between AD patients and controls concerning their differential activation of the dorsal and ventral visual stream. The most pronounced differences were found in the parietal lobule (more activity in controls) and occipito-temporal cortical (more activity in patients). However, the difference between the two groups was

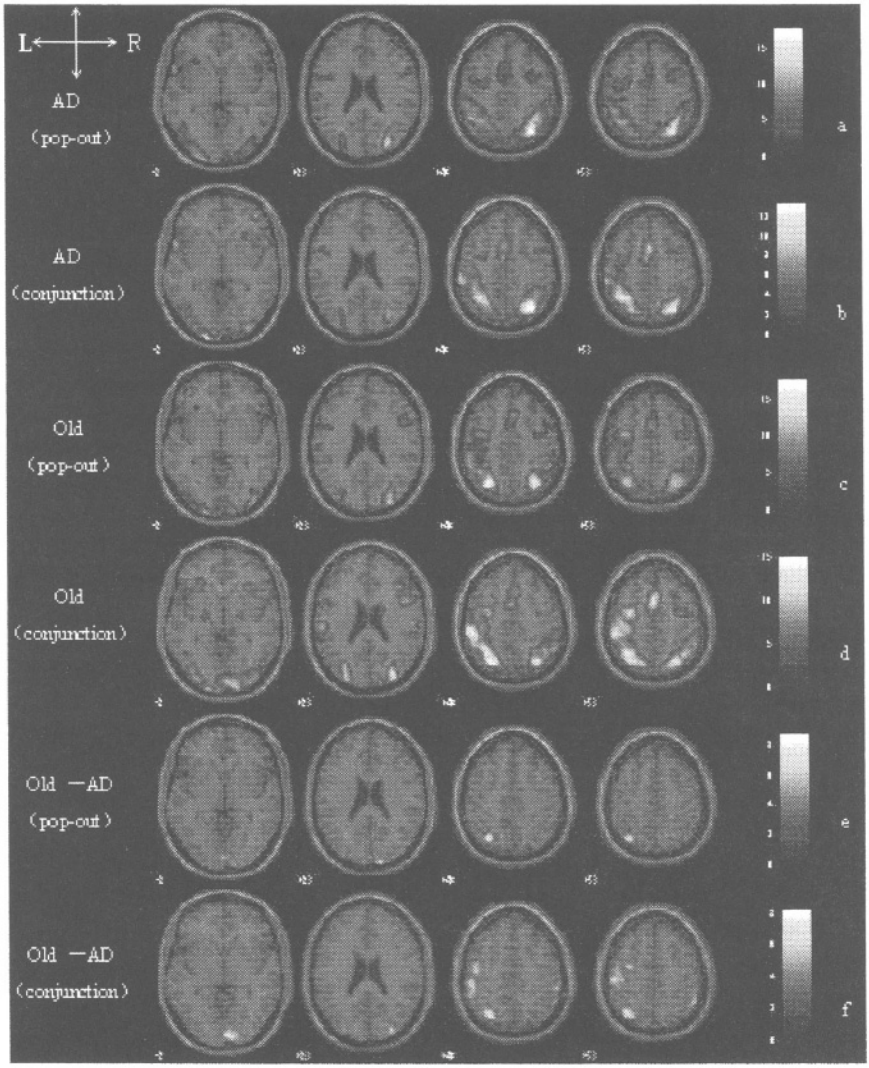


Fig. 2. Averaged brain activation involved in the different conditions (pop-out and conjunction task) of the two groups (AD patients and controls) (a-d) and the comparison between groups for the different conditions (e-f).

small in the pop-out condition. There was less activation in right superior parietal lobule in AD patients, while there was no significant difference in bilateral frontal lobes (Fig. 2). Therefore, AD patients have a particular impairment in the conjunction task but not in the simple-feature task.

4 Discussion

A visual search paradigm composed of simple and conjunction feature search tasks was used to evaluate the level of attentional function in patients with AD and matched controls. This paradigm has previously been used exclusively in the cognitive psychology literature [13,14,15,16]. From behavioral data, We demonstrated that AD patients had significant deficits in visual attention, as revealed by their differentially slowed target detection speed in the conjunction task, and the degree of the impairment was directly related to the size of distrators.

The most important findings from imaging data were that a network related to visual search tasks is very similar in AD patients and controls. The relative contribution of the components of this network differed between the two groups. AD patients showed less activation in the bilateral parietal lobes and right frontal regions, while additional activation was found in left frontal lobes and right occipito-temporal cortical regions in the conjunction task. Given that parietal lobe dysfunction is a typical pathological characteristic of AD, our finding of the impaired conjunction task is consistent with previous work that indicated that the superior parietal cortex is specifically involved in mediating conjunction search [17]. Other possibly relevant brain regions are the anterior cingulated cortex, thought to be involved in selecting target information from distracting information [18] and frontal lobes, thought to be involved in resolving response conflict, both of which may also be abnormal in AD.

Another finding of our study is a double dissociation between patients and controls concerning their differential activation of the dorsal and ventral visual stream. Patients showed significantly less activation in the dorsal stream (SPL), while they revealed higher task-related activity in the right OTC than controls. This shows that in AD, ventral and dorsal visual pathways are not only differently damaged at the input side as demonstrated during passive visual stimulation [19], but these differences remain during active engagement of these regions. Thulborn et al [20] reported reduced parietal cortex activation in the right hemisphere in AD patients during an eye movement task. They interpreted their finding as being a correlate of reduced spatial attention caused by AD. Our results converge with those of Thulborn et al [20]. Probably, this additional remote activation can be interpreted as a potential mechanism to compensate for the reduced functional capacity of the parietal in AD patients.

Interestingly, less resource-demanding capabilities, tapped by the pop-out task, remained relative preserved in AD through the functional compensation of neighboring neural tissues. Of note, it has been suggested that the basal ganglia (which seem to be relatively unaffected in AD) may mediate the fundamental ability to detect salient targets [18].

The results of the present study need to be considered in the context of the large body of neurobiological and neuroimaging study on functional reorganization in AD. Our study was only one step in unraveling the pathophysiology processes of this neurodegenerative disease.

Table 2. Anatomical Regions Activated during the Popout Task ($P < 0.001$)

Age-matched controls			AD patients		
Region	Voxels	X Y Z	Region	Voxels	X Y Z
R-superior parietal lobule(BA7)	18953	32 -66 48	R-superior parietal lobule(BA7)	17202	30 -66 48
L-lingual gyrus(BA18)	18953	-2 -94 -8	R-medial occipital gyrus(BA19)	17202	30 -82 22
R-front eye fields lobule(BA6)	1947	38 -6 60	R-frontal eye fields (BA6)	1916	40 2 58
R-inferior frontal gyrus(BA44)	1947	48 14 28	R-inferior frontal gyrus(BA44)	1916	50 12 32
R-medial frontal gyrus (BA9)	1947	54 12 42	supplementary eye fields(BA6)	1592	2 6 64
Frontal eye fields (BA6)	1295	0 6 54	R-basal ganglia	407	28 18 4
R-superior frontal gyrus(BA8)	1295	6 20 48	R-inferior frontal gyrus(BA45)	407	44 30 16
L-medial frontal gyrus(BA47)	269	-46 50 -6	R-medial frontal gyrus(BA10)	407	40 50 22
L-inferior frontal gyrus(BA47)	269	-28 28 0	L-inferior frontal gyrus(BA47)	249	-26 18 4
R-thalamus	156	-14 -10 14	R-postcentral gyrus	135	64 -18 32

Table 3. Anatomical Regions Activated during the Conjunction Task ($P < 0.001$)

Age-matched controls			AD patients		
Region	Voxels	X Y Z	Region	Voxels	X Y Z
L-superior parietal lobule(BA7)	27717	-32 -56 50	L-superior parietal lobule(BA7)	13813	-36 -42 62
L-precuneus (BA7)	27717	-26 -68 44	L-inferior occipital gyrus(BA18)	13813	-20 -104 0
R-superior parietal lobule(BA7)	8070	32 -60 48	R-superior parietal lobule(BA7)	5280	32 -66 50
L-postcentral gyrus	27717	-50 -32 50	R-medial occipital gyrus(BA19)	5280	34 -94 6
R-frontal eye fields gyrus(BA6)	4030	32 -6 64	R-inferior parietal lobule(BA40)	5280	38 -48 44
R-medial frontal gyrus(BA10)	279	0 6 54	R-medial frontal gyrus(BA10)	1090	44 50 -4
L-basal ganglia	1444	42 56 8	L-medial frontal gyrus(BA46)	476	-48 42 20
R-basal ganglia	1444	18 0 18	L-medial frontal gyrus(BA10)	476	-36 60 12
R-thalamus	1444	14 -6 12	R-inferior temporal gyrus(BA37)	213	44 -64 -14
R-inferior frontal gyrus(BA47)	4030	56 16 2			
R-inferior frontal gyrus(BA46)	279	50 46 14	R-inferior frontal gyrus(BA45)	114	30 26 2

References

1. McKhann,G., Drachman,D., Folstein,M., Katzman,R., Price,D., Stadlan,E.M.: Clinical diagnosis of Alzheimer's disease:Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 34 (1984) 939-944
2. Haxb,J.V., Grady,C.L., Koss,E., Horwitz,B., Heston,L.L., Schapiro,M.B., Friedland, R.P.,Rapoport,S.I.: Longitudinal study of cerebral metabolic asymmetries and associated neuropsychological patterns in early dementia of the Alzheimer type. Archives of Neurology 47 (1990) 753-760
3. Grearing,M., Graves,R.E., Mohr,E.: Breakdown of selective attentional mechanisms in Alzheimer's disease. Journal of Clinical and Experimental Neuropsychology 13 (1991) 69
4. Grande,L., McGlinchey-Berroth,R., Milberg,W.P.,Esposito,M.D.: Facilitation of unattended semantic information in Alzheimer's disease:Evidence from a selective attention task. Neuropsychology 10 (1996) 475-484
5. Nebes,R.D., Brady,C.B.: Focused and divided attention in Alzheimer's disease. Cortex 25 (1989) 305-315
6. Filoteo,J.V., Delis,D.C., Massman,P.J., Demadura,T., Butters,N., Salmon,D.P.: Directed and divided attention in Alzheimer's disease:impairment in shifting of attention to global and local stimuli. J Clin Exp Neuropsychology 14 (1992) 871-883
7. Faust,M.E., Balota,D.A.: Inhibition of return and visuospatial attention in healthy older adults and individuals with dementia of the Alzheimer type. Neuropsychology 11 (1997) 13-29
8. Greenwood,P.M., Parasuraman,R., Alexander,G.E.: Controlling the focus of spatial attention during visual search:effects of advanced aging and Alzheimer disease. Neuropsychology 11 (1997) 3-12
9. Bramer,G.R.: International statistical classification of diseases and related health problems. Tenth revision. World Health Stat.Q 41 (1996) 32-36
10. Folstein,M.F., Folstein,S.E., McHugh,P.R.: Mini-Mental State - A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 12 (1975) 189-198
11. Friston,K.J., Frith,C.D., Liddle,P.F., Frackowiak,R.S.: Comparing functional (PET) images:The assessment of significant change. J. Cereb. Blood Flow Metlab 11 (1991) 690-699
12. Friston,K.J., Holmes,A.P., Poline,J.B., Grasby,P.J., Williams,S.C., Frackowiak,R.S., Turner,R.: Analysis of fMRI time-series revisited. Neuroimage 2 (1995) 45-53
13. Driver,J., McLeod,P., Dienes,Z.: Motion coherence and conjunction search-implications for guided search theory. Perception and Psychophysics 51 (1992) 79-85
14. Pollmann,S.: A pop-out induced extinction-like phenomenon in neurologically int act subjects. Neuropsychologia 34 (1996) 413-425
15. Foster,J.K., Behrmann,M., Stuss,D.T.: Visual attention deficits in Alzheimer's disease:simple versus conjoined feature search. Neuropsychology 13 (1999) 223-245
16. Tales,A., Butler,S.R., Fossey,J., Gilchrist.I.D., Jones,R.W., Troscianko,T.: Visual search in Alzheimer's disease:a deficiency in processing conjunction of features. Neuropsychologia 40 (2002) 1849-1857

17. Corbetta,M., Shulman,G.L., Miezin,F.M., Petersen,S.E.: Superior parietal cortex activation during spatial attention shifts and visual feature conjunction. Science 270 (1995) 802-805
18. Posner,M.I., Driver.J.: The neurobiology of selective attention. Curr. Opin. Neurobiol 2 (1992) 165-169
19. Mentis,M.J., Horwitz,B., Grady,C.L., Alexander.G.E., VanMeter,J., Maisog,J.M., Pietrini,P., Schapiro,M.B., Rapoport,S.I: Visual cortical dysfunction in Alzheimer's disease evaluated with a temporally graded "stress test" during PET. Am. J. Psychiatry 153 (1996) 32-40
20. Thulborn,K.R., Martin,C., Voyvodic,J.T.: Functional MR imaging using a visually guided saccade paradigm for comparing activation patterns in patients with probable Alzheimer's disease and in cognitively able elderly volunteers. Am. J. Neuroradiol 21 (2000) 524-531

Spatio-temporal Identification of Hemodynamics in fMRI: A Data-Driven Approach*

Lirong Yan[1], Dewen Hu[1,2], Zongtan Zhou[1], and Yadong Liu[1]

[1] Department of Automatic Control, National University of Defense Technology, Changsha, Hunan, 410073, P.R.China
[2] Key Lab of Mental Health, Chinese Academy of Sciences, Beijing, 100101, P.R.China
dhu@nudt.edu.cn

Abstract. This paper presents a new data-driven method to identify the spatial and temporal characteristics of the cerebral hemodynamics in functional magnetic resonance imaging (fMRI). The experiments are in block design paradigm and the scans in task blocks are investigated in a sequential manner. Spatial evolvement of the activated regions along with the time-course are demonstrated. The time series of each region is predicated as the convolution of the stimuli with the hemodynamic response function (HRF) formulated as the sum of two gamma functions. The predicted time series is fitted to the actual one by using a nonlinear least-squares procedure to estimate the HRF parameters. Analyses on empirical fMRI datasets exhibit obviously the spatial and temporal dispersion of hemodynamics.

1 Introduction

To date, functional magnetic resonance imaging (fMRI) has been widely used in cognitive neuroscience. When certain tasks are performed, neurons in specific brain areas are activated and their activities trigger the change of blood oxygen level in nearby arterioles, which lays the basis of blood oxygen level-dependent (BOLD) fMRI.

The nature of the BOLD fMRI signal is a consequence of its hemodynamic origins. It is noted that the hemodynamic response (HDR) is much slower than the 10-100 ms timescale of the evoked electrical activity and neurovascular control mechanisms, with reported temporal width on the order of 5-8 s (Dale et al., 1997; Aguirre et al., 1998) and has long time constant undershoots (Glover, 1999). This inherence heavily filters the fMRI signal (Buxton et al., 1997), and therefore hampers the characterization of the actual neuronal response. At the same time, the hemodynamics is a spatial diffused process, so the activated areas detected by HDR are also blurred. The so-called effective temporal and spatial

* Supported by Distinguished Young Scholars Fund of China (60225015), Natural Science Foundation of China (30370416, 60171003), Ministry of Science and Technology of China(2001CCA04100) and Ministry of Education of China (TRAPOYT Project).

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 213–220, 2004.

resolution of BOLD fMRI is physiological and limited primarily by the blurring introduced by the HDR.

It's important to study the characteristic of hemodynamics. Till now most work is based on optical imaging (OI) technique (Malonek et al., 1996; Kim, 2003; Buxton, 2001; Mayhew, 2003) or event-related fMRI paradigms (Buckner et al., 1996; Rosen et al., 1998; Glover, 1999; Miezin, 2000), for OI technique has higher spatial and temporal resolution than fMRI while event-related fMRI designs allow the response to a single stimuli to be examined in a context-independent fashion, hence avoiding the interaction between stimuli which contains nonlinear components (Glover, 1999; Boynton et al., 1996). But, i) OI technique is an invasive technique which can't be applied to normal human, ii) for most of the fMRI analysis a time invariant linear model provides a reasonable approximation to the observed signal (Boynton et al., 1996; Cohen, 1997; Dale et al., 1997; Friston et al., 1998; Glover, 1999) and iii) since 1980's the block design paradigm is popularly applied in most neuroimaging research with easier experimental design and imaging acquisation compared with event-related paradigm.

This paper presents a new method to analyze the fMRI data acquired in a block design experiment, aiming at digging out more useful information of the temporal and spatial dispersion of the HDR.

2 Materials and Methods

BOLD fMRI data were acquired in a GE Signa System operating at 1.5 Tesla with a gradient echo echo-planar imaging (EPI) sequence (TR = 3 sec, TE = 50 ms, FOV=24cm, matrix =64 × 64, slice number = 18, slice thickness = 5 mm, gap = 1.5 mm). Eight healthy, right-handed subjects (four males and four females) participated in the study. These subjects were not pretrained. The experiments were right-hand finger tapping, which occurred in a periodic design containing 5 blocks of 20 scans. Each block consisted of 10 baseline scans followed by 10 scans acquired during task. The whole experiment lasted 300s.

Then we analyzed the data for each subject respectively. Experimental data were imported into the SPM software package (SPM'99, Wellcome Department of Cognitive Neurology). Spatial transformation (realignment, spatial normalization) was performed to correct motion. Data were spatially smoothed with a Gaussian filter (4-mm full-width half-maximum (FWHM) kernel). The latter 4 blocks (80 scans) were included in the further analysis.

2.1 Sequential Activation Detection

In the following processing of parameter estimation and statistical inference, we applied a new data-driven method. The task blocks were analyzed in a sequential manner: except the scans in the control blocks, the first scan in each block after the onset of the stimuli was included while the other nine scans excluded in the time-series analysis and the activation areas were detected; then, the first two scans in each block included and activation areas detected under this condition;

then the first three analyzed and so on, until all the ten scans in each task block were included in activation detection just as normal. The design matrix of GLM in SPM was constructed with a column of box-car functions and a column with 1'sto model the global cerebral blood flow. The contrast was simply (1 0) which indicates 'active > rest'. In this sequential manner, ten groups of results were obtained from one dataset of each subject.

2.2 The Parametric HDR Function

Next, to evaluate the temporal characteristic of the HDR in a parametric way, the hemodynamic response function (HRF, noted as *hrf*) is formulated as the sum of two gamma functions:

Let $f(x,h,l)$ be the probability density function (PDF) of Gamma distribution:

$$f(x) = \frac{l^h x^{(h-1)} \exp(-lx)}{gamma\ (h)} \tag{1}$$

where x is Gamma-variate (Gamma has range $[0,\infty]$), h is the shape parameter $(h > 0)$, l is scale parameter $(l > 0)$. Then the *hrf* can be expressed as:

$$hrf = f(x,h_1,l_1) - f(x,h_2,l_2) * r \tag{2}$$

where h_1, l_1, h_2, l_2 is the shape parameters as (1), r is the parameter which modulates the ratio of peak to undershoot of the HDR. This model of the HDR has been demonstrated reasonable and comprehensive (Friston et al., 1994; Boynton et al., 1996; Martindale et al., 2003; Mayhew et al., 1998) and is utilized in SPM as a kind of basis functions.

The output is predicted by convolving the stimuli with *hrf*. The results of convolution are fitted to the actual time series to determine the five parameters using a nonlinear least-squares procedure (Levenberg-Marquardt algorithm).

3 Results

The activated areas were detected using the *F*-test with $P < 0.001$ (uncorrected). Similar and consistent activations results of the 8 subjects were found in the bilateral premotor and sensorimotor areas, SMA and cerebellum. For our focus is on the spatial and temporal characteristics of HDR, the physiological meaning of the activated areas will not be discussed here. For the convenience, we will present results of one subject as an example.

3.1 Spatial Dispersion of HDR

Ten groups of activation results and their evolvement obtained from one subject are shown in Fig. 1. As can be seen, almost all the areas corresponding to the motor function are detected when only the first scan of task block is entered into

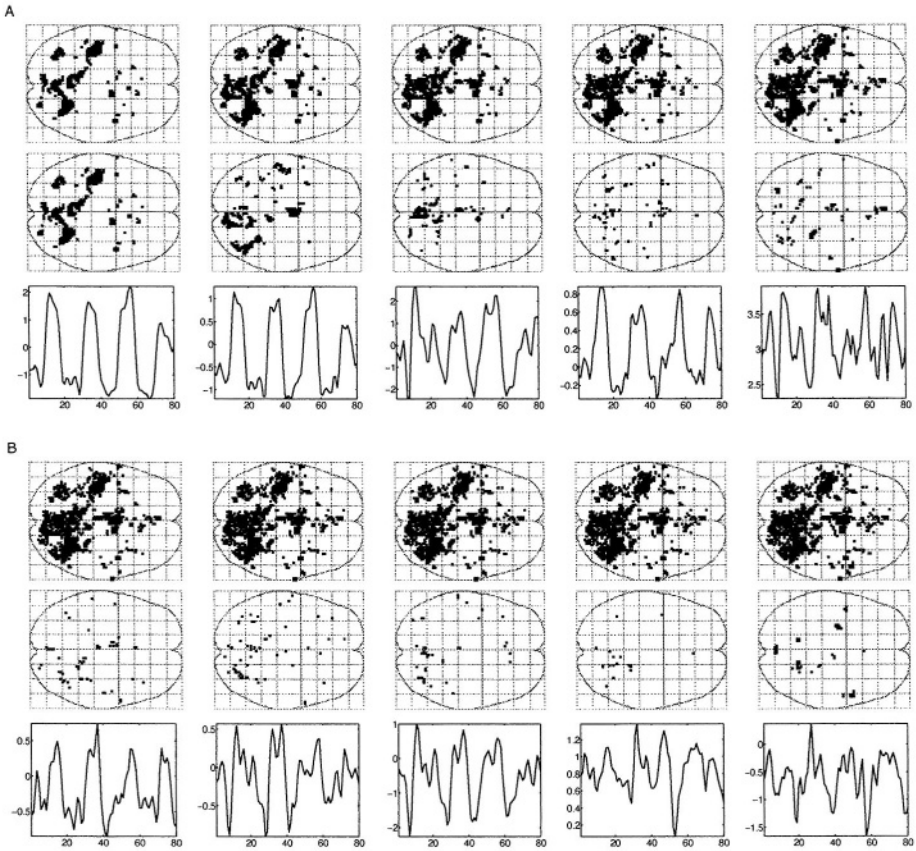


Fig. 1. Results of the sequential method with the number of scans varying from 1 to 5 (A) and 6 to 10 (B). In each group, the first row shows the activation results, the second and third rows show the newly arisen activated areas relative to the former results and their time series (the global cerebral blood flow has been removed).

analysis (the first column of Fig. 1A). With more scans included in the analysis, the newly arisen areas show to be more scattered and their time series are not so obviously periodic and smooth especially when the number of scans is more than five. To make it more obvious, we overlap the first four groups of areas in one map, as shown in Fig. 2. The spatial dispersion of the activated areas, especially in the contralateral sensorimotor areas, cerebellum and SMA, is very remarkable, where the new areas are surrounding or nearby the old areas.

The number of activated voxels (normalized between 0 and 1) in each region of all the eight subjects are shown in Fig. 3. There is a persistent increase until the first seven or eight scans are entered into analysis, while a slight decrease when the first nine or all scans are entered. The initial increase is thought to be a symbol of spatial dispersion of HDR. One may argue that this may be due to

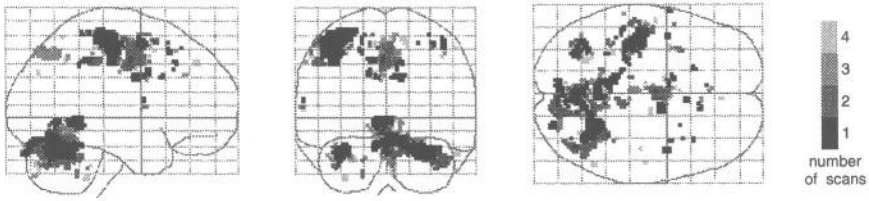


Fig. 2. The first 4 groups of activation results shown in an overlapped mode.

the increasing of statistical power, and we'll discuss this argument and the final decrease later.

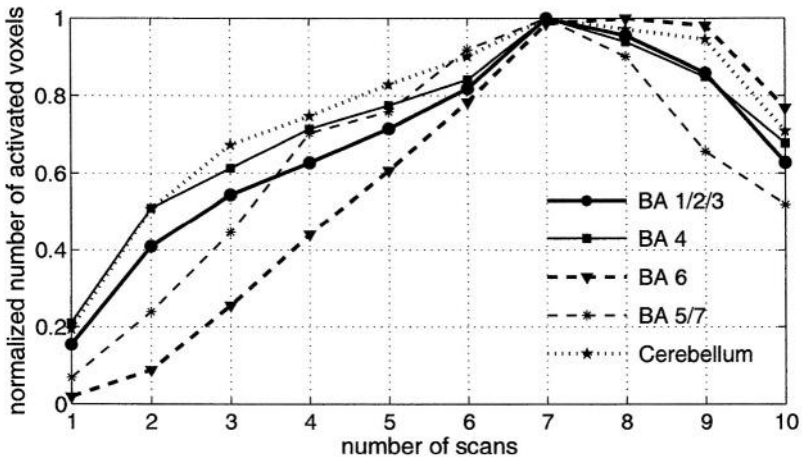


Fig. 3. The number of activated voxels in the main motor areas to the number of scans. Numbers of voxels have been normalized between 0 and 1.

3.2 Temporal Characteristic of HDR

Because the first four scans had higher signal to noise ratio (SNR) (as shown in Fig. 1), the whole time series of areas found active in these scans were extracted to study the temporal characteristic of HDR.

In each group of resultant activated region of interest (ROI), 100 voxels were randomly chosen (if there were less than 100 voxels in the ROI, all voxels in this region were chosen); for each voxel, its whole time series (with the global cerebral blood flow removed) was extracted. The nonlinear least-squares procedure was used to determine the parameters and hence the shape of HRF of each voxel. In Fig. 4, we show the mean HRF of each region. It seems that the temporal difference of the HRF corresponds to different numbers of scans.

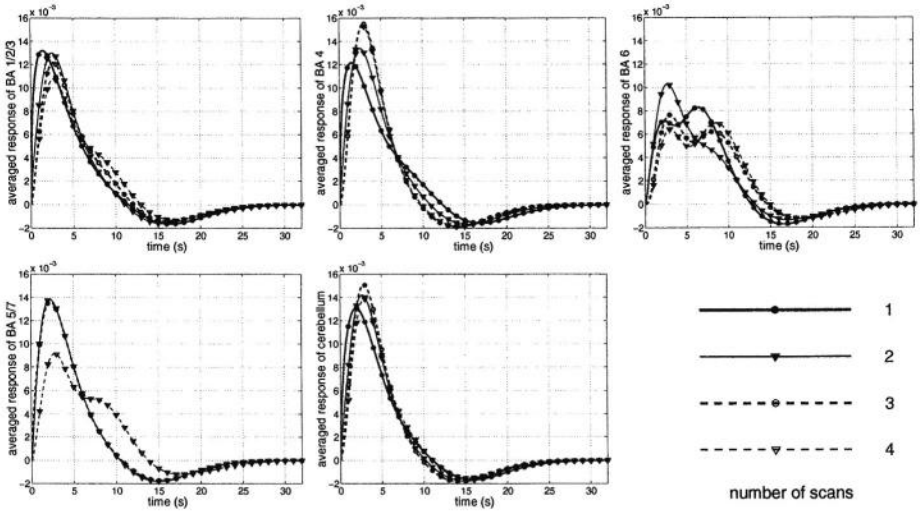


Fig. 4. The mean HDR of different regions and different scans with the number of scans in task block changing from 1 to 4.

To test this hypothesis, in the 100 HRFs in each group of results, time to peak and to undershoot are recorded. Let t_{pi} $(i = 1, 2, 3, 4)$ denote the time to peak when the first i scans are included in analysis. We then test the null hypothesis: $t_{pi} < t_{pj}$ while $i < j$ using the one-tail t-test, the statistic is denoted by T_{ij}. Results are shown in Table 1. It can be seen that if the activated voxels occurred at least two scans apart, the timing difference is always significant (See T_{13}, T_{14}, T_{24} in Table 1); for the adjacent scans, different regions have different characteristics: for cerebellum, all timing difference is significant; for BA 1/2/3, BA 4 and BA 6, difference between t_3 and t_4 is not so significant; for BA 5/7, difference between t_2 and t_3 is not so significant.

Table 1. T level of the compared time to peak of regions with different scan numbers

T level	T_{12}	T_{13}	T_{14}	T_{23}	T_{24}	T_{34}	P_{max}	P_{min}
1/2/3 (C)	-4.946	-6.112	-6.472	-2.116	-2.938	-0.939	0.184	0.000
4(C)	-1.350	-2.343	-2.388	-1.100	-1.140	-0.068	0.472	0.009
6(C)	0.026	-3.028	-3.915	-2.944	-3.787	-0.803	0.402	0.005
5/7(C)	–	–	–	1.849	-6.045	-6.856	0.968	0.000
Cerebellum(I)	-4.593	-7.595	-5.771	-2.408	-3.524	-2.440	0.007	0.000

Note. T_{ij} is the sample statistic for the null hypothesis: $t_{pi} < t_{pj}$ while $i < j$ indicating the number of scans in task block which are included in analysis. P_{max} and P_{min} are respectively the maximum and minimum P-value of the sample statistics.

4 Discussion

The mismatch between blood metabolism and actual neuronal activity makes it very important to study HDR temporally and spatially. As has been referred to, the OI technique and event-related fMRI designs do have advantages in studying the HDR, but the attempt to explore HDR by block design fMRI is also reasonable: i) block design fMRI is more practical and has higher SNR in activation detection than event-related fMRI; ii) by supplying multiple trials, sufficient statistical power and robustness can be obtained. With this understanding a new data-driven method is presented and applied to empirical datasets. Though the results were not so perfect, for the coarse temporal and spatial imaging resolution (TR=3s, voxel size=$3.75 \times 3.75 \times 5mm^3$), some promising results were found. It's believed if the imaging resolution were increased, the estimation of HDR would be more accurate. In fact, EPI method of fMRI can generate images with the temporal resolution of 25ms and spatial resolution of 1mm at most; and the developing fMRI with high field strength can localize neural activity at a columnar level (Thompson et al., 2003). Surely great potential remains in the fMRI experimental design and data analysis.

What is interesting, as has been noted that, the number of activated voxels increases persistent and then decreases with the peak attained at the seventh or eighth scan. The initial increase is due to the spatial dispersion of HDR, not the increasing of statistical power. This could be demonstrated by the significant temporal difference between the voxels activated in different time (Table 1) on the basis that the spatial and temporal characteristics are interrelated and this interrelation would not occur just because of statistical power. The final decrease is illustrated in a statistical framework. For simplicity, we use the model of t-test (F-test is similar as bilateral t-test). As was employed in GLM, the statistical model is:

$$\frac{c^T\hat{\beta} - c^T\beta}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}} \sim t_{J-p} \tag{3}$$

where $c^T\beta$ is the linear compound of the model parameters, $c^T\hat{\beta}$ is the estimated value, $\hat{\sigma}^2$ is the residual sum of squares after the model estimation, X is the design matrix, $J - p$ is the effective degree of freedom of the errors. Generally, with the increase of the degree of freedom, the scope of the region of rejection enlarges, which makes it easier to reject the null hypothesis, that in SPM, there is no activation of a specific voxel. On the other hand, the residual error may not be strictly stationary even with the data smoothed by a Gaussian kernel. As can be seen in Fig. 1, the time series of newly arisen voxels have more high-frequency fluctuation, which may increase the estimated sum of residual squares $\hat{\sigma}^2$, resulting in a decreased sample statistic, hence it is not significant enough to reject the null hypothesis. So the final decrease of the number of activated voxels may be the influence of noise. There seems to be some trade-off in statistical inference. Further investigation would be beneficial to experiment optimization and data analysis.

References

1. Aguirre, G., Zarahn, E., D'Esposito, M.: The variability of human, BOLD hemodynamic responses. NeuroImage 8 (1998) 360–369
2. Boynton, G.M., Engel, S.A., Glover, G.H., Heeger, D.J.: Linear systems analysis of functional magnetic resonance imaging in human V1. J. Neurosci. 16 (1996) 4207–4221
3. Buckner, R.L., Bandettini, P.A., O'Craven, K.M., Savoy, R.L., Petersen, S.E., Raichle, M.E., Rosen, B.R.: Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. Proc. Natl. Acad. Sci. USA 93 (1996) 14878–14883
4. Buxton, R., Frank, L.: A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. J. Cereb. Blood Flow Metab. 17 (1997) 64–72
5. Buxton, R.B.: The elusive initial dip. NeuroImage 13 (2001) 953–958
6. Cohen, J.D., Perlstein, W.M., Braver, T.S., Nystrom, L.E., Noll, D.C., Jonides, J., Smith, E.E.: Temporal dynamics of brain activation during a working memory task [see comments]. Nature 386 (1997) 604–608
7. Dale, A.M., Buckner, R.L.: Selective averaging of rapidly presented individual trials using fMRI. Hum. Brain Mapp. 5 (1997) 329–340
8. Friston, K.J., Jezzard, P., Turner, R.: Analysis of functional MRI time-series. Hum. Brain Mapp. 1 (1994) 153–171
9. Friston, K.J., Josephs, O., Rees, G., Turner, R.: Nonlinear event-related responses in fMRI. Magn. Reson. Med. 39 (1998) 41–52
10. Glover, G.H.: Deconvolution of impulse response in event-related BOLD fMRI. NeuroImage 9 (1999) 416–429
11. Thompson, J.K., Peterson, M.R., Freeman, R.D.: Single-Neuron Activity and Tissue Oxygenation in the Cerebral Cortex. Science 299 (2003) 1070–1072
12. Kim, S.G.: Progress in understanding functional imaging signals. Proc. Natl. Acad. Sci. USA 100 (2003) 3550–3552
13. Malonek, D., Grinvald, A.: Interactions between electrical activity and cortical microcirculation revealed by imaging spectroscopy: Implications for functional brain mapping. Science 272 (1996) 551–554
14. Martindale, J., Mayhew, J., Berwick, J., Jones, M., Martin, C., Johnston, D., Redgrave, P., Zheng, Y.: The hemodynamic impulse response to a single neural event. J. Cereb. Blood Flow Metab. 23 (2003) 546–555
15. Mayhew, J., Hu, D., Zheng, Y., Askew, S., Hou, Y., Berwick, J., Coffey, P., Brown, N.: An evaluation of linear model analysis techniques for processing images of microcirculation activity. NeuroImage 7 (1998) 49–71
16. Mayhew, J.: A measured look at neuronal oxygen consumption. Science 299 (2003) 1023–1024
17. Miezin, F.M., Maccotta, L., Ollinger, J.M., Petersen, S.E., Buckner, R.L.: Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. NeuroImage 17 (2000) 735–759
18. Rosen, B.R., Buckner, R.L., Dale, A.M.: Event-related functional MRI: Past, present, and future. Proc. Natl. Acad. Sci. USA 95 (1998) 773–780

Left Ventricular Motion Estimation Under Stochastic Uncertainties

Huafeng Liu[1,2], Zhenghui Hu[3], and Pengcheng Shi[2]

[1] State Key Laboratory of Modern Optical Instrumentation
Zhejiang University, Hangzhou, China
[2] Department of Electrical and Electronic Engineering
Hong Kong University of Science and Technology, Hong Kong
{eeliuhf@ust.hk, eeship@ust.hk}
[3] Bio-X Laboratory
Zhejiang University, Hangzhou, China
jingke@zju.edu.cn

Abstract. Reliable estimation of regional cardiac deformation is of great importance for the clinical assessment of myocardial viability. Given partial, noisy, image-derived measurements on the cardiac kinematics, prior works on model-based motion estimation have often adopted deterministic constraining models of mathematical or mechanical nature. In this paper, we present a novel estimation framework for motion analysis under stochastic uncertainties. The main novelty is that the statistical properties of the model parameters, system disturbances, and measurement errors are not treated as constant but rather spatio-temporally varying. An expectation-maximization (EM) framework, in both space and time domains, is used to automatically adjust the model and data related matrices in order to better fit a given measurement data set, and thus provides more accurate tissue motion estimates. Physiologically meaningful displacement fields and strain maps have been obtained from *in vivo* cardiac magnetic resonance phase contrast image sequences.

1 Introduction

Noninvasive quantification of heart motion and deformation is of great importance for the diagnosis of coronary artery diseases. With the recent progress in magnetic resonance imaging techniques, particularly MRI tagging and phase contrast velocity imaging, many research and development efforts have been devoted to the image-based regional function analysis of the myocardium [1]. Typically, sparse frame-to-frame correspondences are first established between salient features such as boundary and tagging points. With these noisy kinematics measurements, additional constraining models of mathematical or mechanical nature are then needed to regularize the ill-posed inverse problem and to obtain the dense field motion fields in some optimal senses [1]. Classical frame-to-frame strategies include mathematically motivated spatial regularization and continuum biomechanics based energy minimization while several other developments impose the all important spatiotemporal constraints to perform multi-frame analysis throughout the cardiac cycle [3]

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 221–228, 2004.

So far, it is typically assumed that the mathematical or mechanical constraining models are completely known *a priori,* and are appropriate for the particular image data. For practical situations, especially those pathological cases, however, the constraining models require detailed knowledge about tissue properties which is not only unavailable and is actually the very information one strives to derive from the images. This issue has imposed a significant limitation on the model-based image analysis strategies and their applications, and it has been recognized by a few researchers in relevant areas. An automatic adaptation technique for the elastic parameters of deformable models is proposed within a Kalman filter framework for shape estimation applications [6], where the variation of the elastic parameters depends on the distance of the model from the data and the rate of change of this distance. Similarly, in [8], dynamic finite element refinement is proposed for nonrigid motion tracking, realized through minimizing the error between the actual and predicted behavior. We have also presented several algorithms for simultaneous estimation of tissue motion and elasticity, including the extended Kalman filter [7], the maximum a posteriori (MAP) estimation [4], and the iterative sequential $\mathcal{H}_\infty$ filtering framework [5].

In this paper, we present a robust estimation procedure for cardiac dynamics that is affected by stochastic uncertainties. We differ from previous efforts in two aspects. First, rather than characterizing the model and data noise parameters *explicitly,* we deal with the uncertainties from stochastic state space model point of view, thus allow the introduction of robust filtering method into the nonrigid motion analysis field. Secondly, while recursive minimum mean-square error (MMSE) filtering has been widely adopted in earlier works, it assumes that the disturbances are Gaussian noises with known statistics. In practice, the noise characteristics are typically set from prior experience, and the performance of the MMSE estimators are highly dependent on selection of the disturbance covariance matrices. In our current implementation, an expectation-maximization (EM) estimation procedure is used to derive the best-of-fit between the model/noise parameters and the given measurement data. Constructing the myocardial state space models from biomechanics principles, at each time step, the current estimates of system and noise covariance matrices are used in the expectation (E) step, and the current estimates of the kinematics are used in the maximization (M) step. The EM algorithm therefore involves *maximization* of the *expectation* in an iterative manner until convergence, which will alter the model/noise parameters to better fit the data and thus provide more accurate motion estimation results.

2 Methodology

2.1 Cardiac Dynamics and State Space Representation

By assuming linear material model undergoing infinitesimal deformation, and with finite element representation, the myocardial dynamics equation in terms of the tissue displacement field U takes the form:

$$M\ddot{U} + C\dot{U} + KU = R \tag{1}$$

where R is the system load, M is the mass matrix, K is the stiffness matrix, and $C = \alpha M + \beta K$ is the Rayleigh damping matrix with small α and β for low damping myocardial tissue. The cardiac dynamics can then be converted into a state-space representation:

$$\dot{x}(t) = A_c x(t) + B_c w(t) \tag{2}$$

with:

$$x(t) = \begin{bmatrix} U(t) \\ \dot{U}(t) \end{bmatrix}, \; w(t) = \begin{bmatrix} 0 \\ R(t) \end{bmatrix}, \; A_c = \begin{bmatrix} 0 & I \\ -M^{-1}K & -M^{-1}C \end{bmatrix}, \; B_c = \begin{bmatrix} 0 & 0 \\ 0 & M^{-1} \end{bmatrix}$$

Implicitly assuming hidden Markov process of the system, and including the additive, zero-mean, white process noise $v(t)$ $(E[v(t)] = 0, E[v(t)v(s)'] = Q_s(t)\delta_{ts})$, we can arrive at the discrete dynamics equation:

$$x_{t+1} = Ax_t + Bw_t + v_t \tag{3}$$

with $A = e^{A_c T}$ and $B = A_c^{-1}(e^{A_c T} - I)B_c$, where T is the image frame interval.

An associated measurement equation, which describes the observed imaging data $y(t)$, takes the following form:

$$y_t = Dx_t + e_t \tag{4}$$

where D is the known measurement matrix that is constructed dependent on the types of data inputs, and $e(t)$ is the measurement noise which is additive, zero mean, and white $(E[e(t)] = 0, E[e(t)e(s)'] = R_o(t)\delta_{ts}$, independent of $v(t))$.

2.2 The EM Estimation Strategy

Eqns. 3 and 4 represent a continuous-time system with discrete-time measurements, or a so-called sampled data system, which describe the cardiac dynamics and the image-derived noisy and partial kinematics measurements. And the aim is then to recover the complete motion/deformation fields.

Since it is still difficult to reliably measure the material-specific Young's modulus and Poisson's ratio *in vivo,* the system matrices $\{A, B\}$ and the noise covariance matrices $\{Q_s, R_o\}$ in the state equations are actually not known *a priori,* at least not accurately. For example, if the actual system matrices were actually $\{A + \delta A, B + \delta B\}$, until now, all aforementioned works are based on $\{A,B\}$ alone, without accounting for the existence of $\{\delta A, \delta B\}$. And this inexactness seriously effects the accuracy and reliability of the estimation results. Here, we will introduce a expectation maximization framework for state-space estimation which accounts for the statistical variations of the system and noise matrices.

Basic Concepts The initial state x_1 distribution is assumed Gaussian white with mean $\hat{x}_1$ and covariance Σ_1:

$$p(x_1) = \frac{\exp\{-\frac{1}{2}[x_1 - \hat{x}_1]^T Q_s^{-1}[x_1 - \hat{x}_1]\}}{2\pi^{\frac{k}{2}}|\Sigma_1|^{\frac{1}{2}}} \tag{5}$$

Due to the properties of Gaussian, all future states and observations are also Gaussian distributed. The conditional probability distribution (PDF) of the next state can be written iteratively as:

$$p(x_t|x_{t-1})) = \frac{\exp\{-\frac{1}{2}[x_t - Ax_{t-1} - Bw_{t-1}]^T Q_s^{-1}[x_t - Ax_{t-1} - Bw_{t-1}]\}}{2\pi^{\frac{k}{2}}|Q_s|^{\frac{1}{2}}} \quad (6)$$

and the conditional PDF of $y(t)$ given $x(t)$ is:

$$p(y_t|x_t) = \frac{\exp\{-\frac{1}{2}[y_t - Dx_t]^T R_o^{-1}[y_t - Dx_t]\}}{2\pi^{\frac{p}{2}}|R_o|^{\frac{1}{2}}} \quad (7)$$

where $|Q_s|$ and $|R_o|$ are the determinants of matrix Q_s and R_o respectively.

Likelihood Function Denote a sequence of kinematics observations $y_{1:N} = [y_1, ..., y_N]$ and system states $x_{1:N} = [x_1, ..., x_N]$, where N is the total number of image frames, x_t is the state of all the myocardial points (fixed number) at frame t, and y_t all the observations (varying size due to the time-dependent imaging data) at frame t. By the Markov property implied by our state model, the joint likelihood for $x_{1:N}$ and $y_{1:N}$ is:

$$p(x_{1:N}, y_{1:N}) = \prod_{t=1}^{N} p(y_t|x_t) \prod_{t=2}^{N} p(x_t|x_{t-1}) p(x_1) \quad (8)$$

Therefore, the joint log probability is a sum of quadratic terms:

$$\begin{aligned} lnp(x_{1:N}, y_{1:N}) = & -\sum_{t=1}^{N}(\frac{1}{2}[y_t - Dx_t]^T R_o^{-1}[y_t - Dx_t]) \\ & -\sum_{t=2}^{N}(\frac{1}{2}[x_t - Ax_{t-1} - Bw_{t-1}]^T Q_s^{-1}[x_t - Ax_{t-1} - Bw_{t-1}]) \\ & -\frac{1}{2}[x_1 - \hat{x}_1]^T \Sigma_0^{-1}[x_1 - \hat{x}_1] - \frac{N}{2}\ln|R_o| \\ & -\frac{N-1}{2}\ln|Q_s| - \frac{1}{2}\ln|\Sigma_0| - \frac{N(p+k)}{2} ln2\pi \end{aligned} \quad (9)$$

The EM Algorithm The EM algorithm consists of two sequential steps: the E-step which calculates the expectation of the log likelihood, termed the $\mathcal{Q}$-function, and the M-step which seeks to maximize the $\mathcal{Q}$-function with respect to some parameter vector Θ in order to generate a new estimate Θ^{new}. In our framework, the parameter vector is $\Theta = \{A, B, Q_s, R_o\}$.

Conceptually, for the E-step, with all the parameters A, B, Q_s, and R_o fixed, one estimates the state variable x; then for the M-step, with the value of the state variable x known (from the E-step), one updates the parameter vector Θ. This process is iteratively executed until both x and Θ reach convergence.

E-Step: The E-step requires the computation of the expected log likelihood given the observation $y_{1:N}$ and parameter vector $\Theta = \{A, B, Q_s, R_o\}$ estimated at the previous step, that is:

$$\mathcal{Q} = E[lnp(x_{1:N}, y_{1:N})|y_{1:N}] \tag{10}$$

Denoting $x_{t|\tau} = E[x_t|y_{1:\tau}]$ and $V_{t|\tau} = Var[x_t|y_{1:\tau}]$, evaluation of the $\mathcal{Q}$-function requires the computation of the following three expectations:

$$x_{t|N} = E[x_t|y_{1:N}] \tag{11}$$

$$P_{t|N} = E[x_t x_t'|y_{1:N}] = V_{t|N} + x_{t|N} x_{t|N}' \tag{12}$$

$$P_{t,t-1|N} = E[x_t x_{t-1}'|y_{1:N}] = V_{t,t-1|N} + x_{t|N} x_{t-1|N}' \tag{13}$$

with $P_{t|N}$ and $P_{t,t-1|N}$ referred to as the state correlations, and $V_{t|N}$ and $V_{t,t-1|N}$ as the state error correlations. The three expectations in (11) to (13) can be readily obtained by running a Kalman smoother (KS), thus allowing the $\mathcal{Q}$-function to be computed. The standard Kalman smoother can be realized through iterative steps of a forward extended Kalman filter (KF) followed by a backward smoothing filter.

The operation of the linear Kalman filter adopts a form of feedback control in estimation: the filter estimates the process state at some time and then obtains feedback in the form of measurements. As such, the equations for the KF fall into two groups: time update equations and measurement update equations. The time update equations are responsible for projecting forward the current state and error covariance estimates to obtain the *a priori* estimates for the next time step, while the measurement update equations are responsible for the feedback. With initialization $x_{1|0} = \hat{x}_1$ and $V_{1|0} = \Sigma_1$, the state estimates and the error covariance matrices are computed sequentially:

1. Time update equations, *the predictions,* for the state

$$x_{t|t-1} = Ax_{t-1|t-1} + Bw_{t-1} \tag{14}$$

and for the error covariance

$$V_{t|t-1} = AV_{t-1|t-1}A' + Q_s \tag{15}$$

2. Measurement update equations, *the corrections,* for the Kalman gain

$$L_t = V_{t|t-1}D'(DV_{t|t-1}D' + R_o)^{-1} \tag{16}$$

for the state

$$x_{t|t} = x_{t|t-1} + L_t(y_t - Dx_{t|t-1}) \tag{17}$$

and for the error covariance

$$V_{t|t} = V_{t|t-1} - L_t(DV_{t|t-1}D' + R_o)L_t' \tag{18}$$

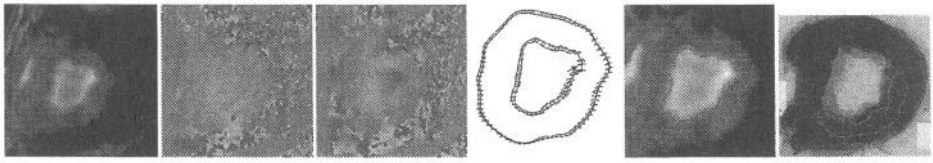


Fig. 1. From left to right: the magnitude, x-velocity, and y-velocity MR phase contrast images; the boundary displacement and mid-wall velocity constraints (system inputs); and the post mortem TTC staining of the myocardium, with infarcted zone highlighted.

Further, the backward smoothing part of the KS algorithm can be achieved through the well-known Rauch-Tung-Striebel smoother [2],:

$$\Lambda_{t-1} = V_{t-1|t-1}A'(V_{t|t-1})^{-1} \tag{19}$$

$$x_{t-1|N} = x_{t-1|t-1} + \Lambda_{t-1}(x_{t|N} - Ax_{t-1|t-1}) \tag{20}$$

$$V_{t-1|N} = V_{t-1|t-1} + \Lambda_{t-1}(V_{t|N} - V_{t|t-1})\Lambda'_{t-1} \tag{21}$$

$$P_{t|N} = V_{t|N} + x_{t|N}x'_{t|N} \tag{22}$$

$$V_{t-1,t-2|N} = V_{t-1|t-1}\Lambda'_{t-2} + \Lambda_{t-1}(V_{t,t-1|N} - AV_{t-1|t-1})\Lambda'_{t-2} \tag{23}$$

$$P_{t,t-1|N} = V_{t,t-1|N} + x_{t|N}x'_{t-1|N} \tag{24}$$

With these KS equations, it is thus possible to derive the $\mathcal{Q}$-function.

M-Step: In the M-step, the $\mathcal{Q}$ function is maximized with respect to the parameter set $\Theta = \{A, B, Q_s, R_o\}$. This can be achieved by taking the corresponding partial derivative of the expected log likelihood, setting to zero, and computing the following:

$$[A^{new}, B^{new}] = \sum_{t=2}^{N} \begin{bmatrix} V_{t,t-1|N} + x_{t|N}x'_{t-1|N} & x_{t|N}w'_{t-1} \end{bmatrix} \left\{ \sum_{t=2}^{N} \begin{bmatrix} V_{t-1|N} + x_{t-1|N}x'_{t-1|N} & x_{t-1|N}w'_{t-1} \\ w_{t-1}x'_{t-1|N} & w_{t-1}w'_{t-1} \end{bmatrix} \right\}^{-1} \tag{25}$$

$$Q_s^{new} = \frac{1}{N-1} E\left[\sum_{t=1}^{N-1} v'_t v_t | y_{1:N} \right] \tag{26}$$

$$R_o^{new} = \frac{1}{N} \sum_{t=1}^{N} (y_t y'_t - Dx_{t|N}y'_t) \tag{27}$$

3 Results and Discussions

The EM framework is used to estimated the kinematics parameters of the canine left ventricle described by the cine MR phase contrast image sequence

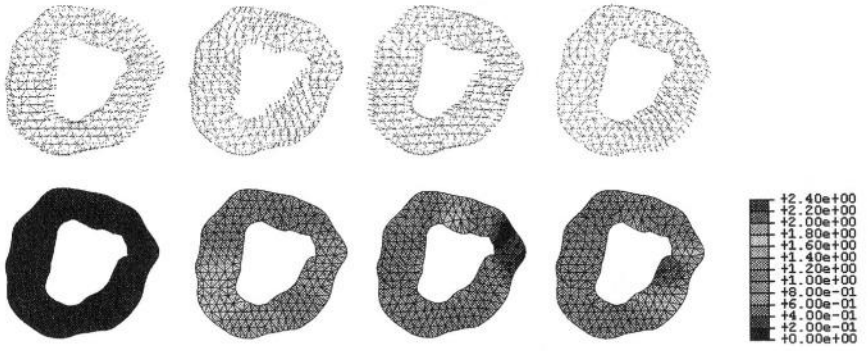


Fig. 2. Estimated frame-to-frame displacement direction (top) and cumulative displacement magnitude (w.r.t. frame #1) maps (bottom): frames #1, #5, #9, and #13.

(Fig. 1). Also shown is the highlighted regional volume of the myocardial injury zone, which is found from the triphenyl tetrazolium chloride (TTC) stained post mortem myocardium and provides the clinical gold standard for the assessment of the image analysis results. Myocardial boundaries and frame-to-frame boundary displacements are extracted using an active region model strategy [9], and the inputs to the system include these boundary displacements and the instantaneous tissue velocity information from the phase contrast images.

Estimated displacement and strain maps are shown for selected image frames in Fig. 2 and Fig. 3 respectively. From the displacement maps, little contracting motion is observed at the infarct zone (lower right quarter) during the contraction phase of the cardiac cycle (i.e. frames #1 to #4), which is being pushed out by increased ventricular pressure generated by the active contraction of the healthy tissue. The injury zone starts to contract when the normal tissue stop its contraction (frame #5), and continues while the normal tissue starts to expand (frames #9 to #10). The expansion at the infarct zone does not occur until frame #13. Similar observations can be made from the radial (R), circumferential (C), and R-C strain maps, where it is easy to notice that the infarct region has substantially different deformation characteristics from the other tissues: little deformation in the radial and circumferential directions at all phases of cardiac cycle, while substantial deformation changes in the shear strain maps.

From the displacement and strain maps, there are consistent signs of myocardial impairment at the lower right quarter of the LV, which agree very well with TTC staining. This indicates the possibility of using this noninvasive, image-based motion analysis technique for the diagnosis of ischemic heart diseases.

Acknowledgments Thanks to Dr. Albert Sinusas of Yale University for the canine experiments and the imaging data. This work is supported by HKRGC-CERG HKUST6151/03E and National Basic Research Program of China (No: 2003CB716104).

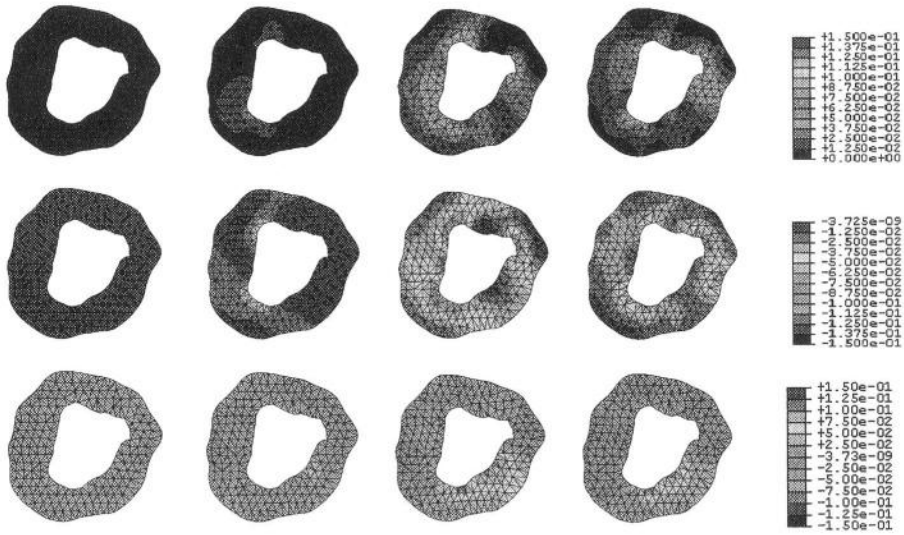


Fig. 3. Estimated displacement direction (top) and magnitude maps (bottom), w.r.t end-diastole (frame #1): frames #1, #5, #9, and #13.

References

[1] A.J. Frangi, W.J. Niessen, and M.A. Viergever. Three-dimensional modeling for functional analysis of cardiac images: A review. *IEEE Transactions on Medical Imaging,* 20(1):2–25, 2001.

[2] A. Gelb. *Applied Optimal Estimation.* MIT Press, Cambridge, MA, 1974.

[3] W.S. Kerwin and J.L. Prince. The kriging update model and recursive space-time function estimation. *IEEE Transactions on Signal Processing,* 47(11):2942–2952, 1999.

[4] H. Liu and P. Shi. Simultaneous estimation of left ventricular motion and material properties with maximum a posteriori strategy. In *IEEE Computer Vision and Pattern Recognition,* pages 161–169, 2003.

[5] H. Liu and P. Shi. Cardiac motion and elasticity characterization with iterative sequential $\mathcal{H}_\infty$ criteria. In *Medical Image Computing and Computer Assisted Intervention,* page in press, 2004.

[6] D.N Metaxas and I.A. Kakadiaris. Elastically adaptive deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 24(10):1310–1321, 2002.

[7] P. Shi and H. Liu. Stochastic finite element framework for simultaneous estimation of cardiac kinematics functions and material parameters. *Medical Image Analysis,* 7(4):445–464, 2003.

[8] L.V. Tsap, D.B. Goldof, and S. Sarkar. Nonrigid motion analysis based on dynamic refinement of finite element models. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 22(5):526–543, 2000.

[9] A.L.N. Wong, H. Liu, A. Sinusas, and P. Shi. Spatiotemporal active region model for simultaneous segmentation and motion estimation of the whole heart. In *IEEE Workshop on Variational, Geometric and Level Set Methods in Computer Vision,* pages 193–200, 2003.

Combined CFD/MRI Analysis of Left Ventricular Flow

R. Merrifield[1], Q. Long[2,*], X.Y. Xu[1], P.J. Kilner[3], D.N. Firmin[3], G.Z. Yang[1]

[1]Faculty of Engineering, Imperial College, London, SW7 2BZ, UK
[2] Brunel Institute for Bioengineering, Brunel University,
Uxbridge, Middlesex, UB8 3PH, UK
quan.long@brunel.ac.uk
[3] Cardiovascular MR Unit, Royal Brompton Hospital
Imperial College, London, SW7 2BZ, UK

Abstract: The relationship between the morphology and blood flow of the Left Ventricle (LV) during myocardial remodelling is complex and not yet fully understood. Cardiovascular MR (CMR) velocity imaging is a versatile tool for the observation of general flow patterns *in-vivo*. More detailed understanding of the coupled relationship between blood flow patterns and myocardial wall motion can be further enhanced by the combined use of Computational Fluid Dynamics (CFD) and CMR. This permits the generation of comprehensive high-resolution velocity fields and the assessment of dynamic indices, such as mass transport and wall shear stress, that are important but cannot be measured directly by using imaging alone. One of the key drawbacks of ventricular flow simulation using CFD is that it is sensitive to the prescribed inflow boundary conditions. Current research in this area is limited and the extent to which this affects *in-vivo* flow simulation is unknown. In this work, we measure this sensitivity as a function of the inflow direction and determine the limit that is required for accurate ventricular flow simulation. This represents an important step towards the development of a combined MR/CFD technique for detailed LV flow analysis.

Keywords: Cardiovascular Magnetic Resonance, Computational Fluid Dynamics, Left Ventricle Flow. Boundary Condition.

1 Introduction

Heart disease is one of the biggest killers and debilitating factors in the world. Coronary atherosclerosis leading to myocardial infarction can be immediately fatal in almost a third of the patients involved. For the survivors, heart failure may follow, which carries a poor prognosis despite improvements in treatment with modern techniques. In cases where a heart attack is not immediately fatal, the shape and function of the ventricles can change over the following weeks and months. This

* Corresponding Author

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 229-236, 2004.

natural process of myocardial remodelling first involves the expansion of the infarcted myocardium, but can continue to affect the adjacent healthy tissue until the overall structure of the heart is altered. As the condition progresses, the functionality of the heart can deteriorate. The extent to which this happens can vary considerably between patients. Understanding the process of remodelling and its relationship with cardiac function is vital to the understanding of heart failure and the subsequent morbidity and mortality associated with myocardial infarction.

In general, LV dysfunction involves a number of interrelated events both in systole and diastole. Each of these factors, including ventricular relaxation, diastolic filling, ventricular contraction, pericardial restraint and ventricular interaction, is interrelated to the others in a complex sequence of events. A detailed investigation of intra-ventricular flow patterns could provide practical insight to the different relationships involved and facilitate the diagnosis and treatment of LV dysfunction. Thus far, no single measurement technique is available to offer detailed quantitative information about the time-dependent LV flow patterns and the impact of ventricular movements on haemodynamic changes. Catheter-based techniques are invasive and not practical for widespread application or serial follow-up examinations, whereas Doppler flow imaging is practically limited by the angle between the flow and the ultrasonic beam. MR phase contrast velocity mapping is perhaps the most versatile technique for *in-vivo* flow measurement but prolonged data acquisition time and limited spatio-temporal resolution are major limiting factors. In parallel to the development of non-invasive imaging techniques, CFD has made significant improvements in recent years. It has been proven to be an effective means of studying complex cardiovascular dynamics, being able to provide detailed haemodynamic information that is unobtainable using direct imaging techniques. CFD is concerned with the numerical solution of a set of partial differential equations that govern the flow. In practice, the discretized equations are solved by using numerical approaches within a computational mesh. It is necessary to specify a set of boundary conditions at the inflow and outflow regions of the flow domain. Once a solution has been reached, the CFD technique is able to ascertain the velocity and pressure at each grid and time point within the mesh.

Existing research has so far utilised CFD techniques to simulate blood flow in the heart with varying degrees of realism. The early computational models developed in the 70-80's were confined to 1D or 2D, aimed at examining the global flow patterns, pressure waveforms and transmitral flow in simplified geometries [1],[2]. These models were subsequently improved to incorporate more realistic geometries and fluid/ventricular wall interactions in order to obtain velocity and pressure patterns in the ventricles, as well as stress distributions within the wall [3-12].

To investigate the sensitivity of the prescribed inflow direction to the CFD results, a detailed study is performed in this paper with additional flow information that is directly acquired at the mitral valve using MR phase contrast velocity mapping. A set of 10 simulations were performed by using a model of a healthy left ventricle. Each of these simulations utilised different inflow directions. The differences between the derived flow fields were assessed both qualitatively and quantitatively with that of the MR results. A further set of simulations were then performed to test the consistency of these results across 6 normal subjects. For each subject, a set of 5 different inflow directions were simulated.

2 Methods

2.1 MR Data Acqusition

Imaging was performed on a Siemens Sonata 1.5T MR system. A multi-slice cine True-FISP imaging sequence was used to acquire 7 short axis slices at 16 phases providing complete spatial and temporal coverage of the LV. Each slice was acquired in a single 20 second breath-hold so as to minimize registration errors caused by respiratory movement. MR flow imaging was used to aid the prescription of the inflow boundary conditions and to validate the CFD simulations. To this end, a phase contrast velocity mapping sequence was used to acquire two long-axis images. The first long-axis image was oriented to pass through both the inflow and outflow tracts whereas the second was set to be in the orthogonal direction passing through the inflow tract. For each acquisition, all three orthogonal velocity components of velocity were obtained. Due to the length of the imaging time required, the three velocity components were acquired within separate breath-holds. As with the morphological image acquisition, retrospective cardiac gating was used to specify the acquisition of 20 phases across the cardiac cycle. The average inflow velocity was measured for all phases during the filling of the left ventricle. An elliptical region of $3cm^2$ was delineated in each phase of the cardiac cycle. This region was located just within the left ventricle and adjacent to the mitral valve plane. Care was taken so that the regions only contained blood but not flow artefacts or the valve leaflets. The average of each velocity component was then measured to prescribe the inflow boundary conditions for the ventricular flow simulation.

2.2 Ventricular Modelling

The ventricular models utilised two surface meshes to represent the endocardial border of the LV. The first of these meshes delineated the inflow tract and the main body of the ventricle, whereas the second represented the outflow tract. The cavity of the LV was considered to be the Boolean union of the volumes enclosed by the meshes. Points within the meshes that lay beyond the mitral or aortic valve planes were discarded so that the extent of the flow domain was limited to the ventricular blood pool. To allow for an accurate CFD simulation, it was necessary to prescribe the wall movement of the LV at a higher temporal resolution than that could be acquired using the CMR imaging technique. For this purpose, Catmull-Rohm spline interpolation was used to create intermediate meshes, resulting in a total of 49 time frames over the cardiac cycle. The volume mesh generation scheme was complicated by the incorporation of the dynamic valve planes. For some subjects, it was necessary to raise the position of the valve plane by 1cm in the direction of the left atrium. This ensured that none of the constituent cells demonstrated a negative volume or underwent excessive deformation.

2.3 CFD Simulation

The Navier-Stokes equations for 3D time-dependent laminar flow with moving walls were solved using a finite-volume based CFD solver CFX4 (CFX international, AEA technology, Harwell). The blood was treated as an incompressible Newtonian fluid with a constant viscosity of 0.004 Kg/(ms). The simulation started from the beginning of systole with zero pressure defined at the aortic valve plane while the mitral valve plane was treated as non-slip wall. At the beginning of diastole, the aortic valve was closed by treating it as a wall, whilst the mitral valve plane was opened by using a combination of pressure and velocity boundaries[13]. A plug velocity profile was assumed at the flow boundaries. The inflow direction was determined by the mean transmitral velocity vector obtained from the MR velocity measurements. The simulation was repeated for four cycles to reach a periodic solution. The results obtained in the fourth cycle are presented here.

The first experiment measured the sensitivity of the CFD simulation to the direction of the inflow for a single healthy subject. Firstly, a simulation was performed by using the average inflow direction measured from the MR velocity images. A further 4 simulations were then performed with the inflow direction modified by 5 degrees in each of the 4 orthogonal directions *A', P', I'* and *S'* as defined in the figure 2 (left graph, second row). These 4 simulations were then repeated with the original inflow direction modified by 10 degrees. In order to indicate the sensitivity of simulation at even greater angles, two further simulations were performed with angles of 15 and 20 degrees in the direction *A'*.

The second experiment evaluated the sensitivity of the technique across a further 5 normal subjects. This was designed to examine the reproducibility of the results obtained from the first experiment. A total of 5 simulations were performed for each subject. The first of these utilised the average inflow direction measured from the MR velocity images. This was then followed by simulations with 5 degrees of variation in each of the 4 orthogonal directions.

2.4 Validation

The sensitivity of the simulation process was evaluated by comparing the flow fields generated using the measured inflow direction to each of those generated using a modified inflow direction. To provide quantitative comparison, two parameters that characterised the difference between pairs of 3D flow fields were calculated. The parameters were based on the mean variation in direction and magnitude between their constituent velocity vectors. It was investigated how these parameters varied as a function of the prescribed inflow direction. This gave an indication of the precision with which the boundary conditions should be specified in order to perform reproducible ventricular flow simulations.

3 Results

Figure 1 demonstrates the typical correspondences between the simulated ventricular flow and that measured by using MR imaging. Two orthogonal planes are presented for each of the techniques which are anterior-posterior (A'-P') and Inferior-superior (I'-S') as defined in figure 2 (left graph, second row). The length of each arrow is proportional to the magnitude of the in-plane velocity. It can be seen that the flow fields have a similar overall topological structure but regional differences are evident. Both flow fields consist of an inflow jet directed into the expanding ventricular cavity. The direction of this jet is the same for both techniques as the measured inflow velocity is used to prescribe the boundary conditions. The discrepancy in absolute inflow velocity is due to the fact that only the inflow direction rather than the absolute value is constrained. Overall, the flow fields measured by using MR velocity imaging had a higher complexity than those derived by simulation. For example, regions containing small vortices were often present in the measured flow fields but not in the simulated results. These regions were typically located adjacent to the inflow tract and around the papillary muscles. This is not surprising as detailed anatomical structure in these regions was not modelled.

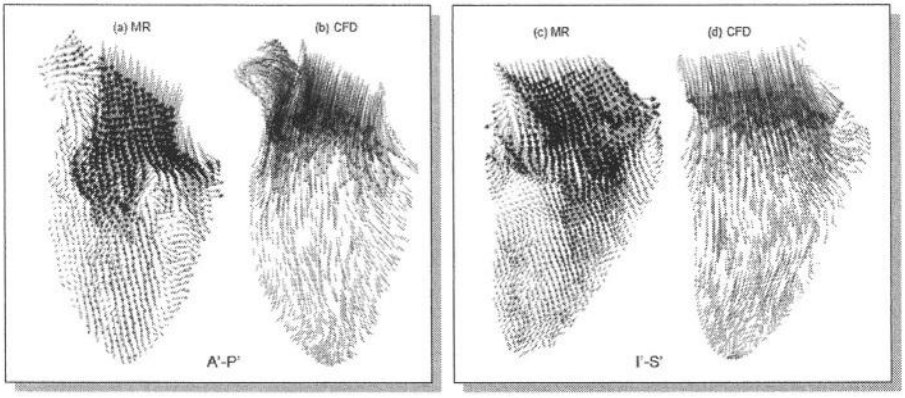


Figure 1: A comparison of flow fields obtained by MR velocity imaging and CFD simulation. For both techniques, two cross sectional velocity maps are displayed to characterise the 3D nature of the flow. It is evident that, although the CFD simulations are not fully realistic, they can produce flow fields with a similar overall topology as those measured by *in-vivo* imaging.

Two parameters were calculated to quantitatively characterise the differences between pairs of flow fields generated by CFD simulation. The first of these was defined to be the ratio of the velocity magnitudes for corresponding points within each flow field. This value was averaged for all points within the ventricular volume. It has been displayed as a function of the cardiac phase within the first row of Figure 2. The cardiac phase has been normalised so that the value 0.0 corresponds to the opening of the mitral valve leaflets and 1.0 represents end-diastole. The second parameter, which is demonstrated in the second column, was defined as the mean angle between velocity vectors at corresponding points within each flow field. It is evident that, in general, the values of both parameters increase throughout each simulation. This is to

be expected due to the lower average flow velocity at later diastole. Of greater importance however, it is demonstrated that the values of both of the parameters increase as a function of the inflow angle. This dependency characterises the sensitivity of the simulation to the inflow angle. Figure 3 shows the consistency of the simulations over 6 normal subjects. Each graph characterises the differences between the simulations performed with the measured inflow direction and those performed with a modified inflow direction. For the figures in this paper, the differences between the 3D flow fields are represented by the mean percentage difference in the velocity magnitudes. It can be seen that, for all subjects, the measured differences between the flow simulations lie within a tight range of 5% to 22% and the simulation were most sensitive to changes in the direction A'.

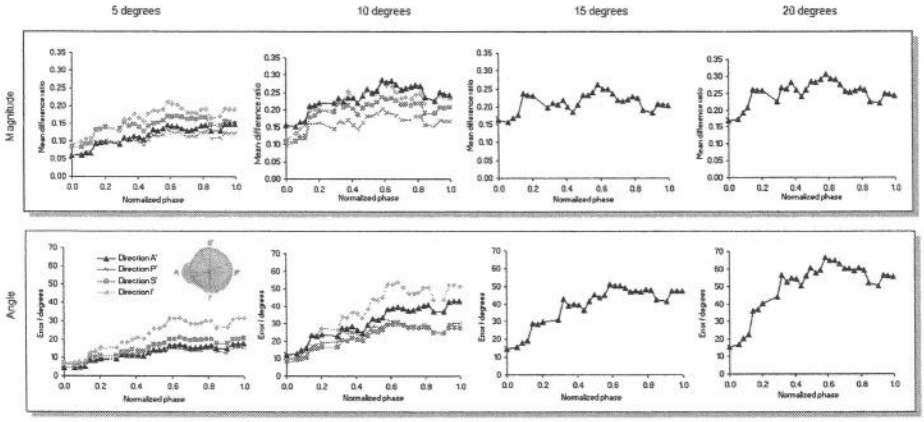


Figure 2: The variability of flow fields generated by CFD simulation. Each graph shows the differences between the flow fields that were generated using the measured inflow direction and those that utilised a modified inflow direction. The two rows show the variation in the mean difference ratio in the velocity magnitude and the mean difference in the angles between corresponding velocity vectors. The orientation parameters A', P', I' S' are defined in the left graph second row.

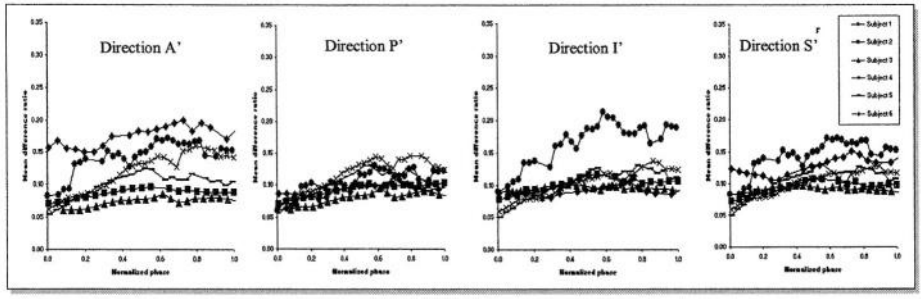


Figure 3: A demonstration of the sensitivity of the flow simulations to the inflow direction for 6 normal subjects (six lines in each graphs). The four graphs represent the different ways in which the inflow direction was modified. x-axis and y-axis correspond to normalised time and mean difference ratio respectively. It can be seen that, in general, the simulations were most sensitive to changes in the direction A'.

4 Discussion and Conclusions

This study has evaluated the sensitivity of ventricular flow simulations to changes in the inflow direction. The aim of this work was to establish how accurately the inflow boundary conditions must be specified in order to give reproducible simulations. It has been shown that changes to the inflow direction of 5 degrees do not significantly affect the flow topology. These changes do bring about slight differences between the magnitudes and directions of the corresponding velocity vectors however. Changes of 10 degrees or more did produce flow patterns with an altered topology. The differences between the velocity vectors were also substantial. It is therefore concluded that, if the topology of the flow is to be assessed, it is necessary to specify the inflow direction within an accuracy of 5 degrees. For detailed quantitative assessment, however, the inflow direction needs to be specified with as high an accuracy as possible. This is a significant finding in that it suggests that for qualitative flow pattern analysis, some of the rapid imaging techniques such as 3D echo combined with Doppler velocity mapping could be used to provide adequate boundary conditions for patient specific flow simulation. The finding also justifies the importance of using MR velocity mapping for providing detailed inflow boundary conditions for accurate quantitative analysis.

In the current study, the plug profile specified for the ventricular inflow is a simplification of the complex flow through the mitral valve. Although it is beyond the scope of the current study to prescribe this profile to a greater accuracy, it would be possible to acquire the relevant data using MR velocity imaging. Two different imaging schemes could be utilised to produce detailed 2D velocity profiles of the inflow region. Firstly, a set of 1D profiles could be acquired from different long-axis images and then interpolated in a radial direction to produce a complete 2D profile. Alternatively, a more sophisticated imaging sequence could be developed to track the mitral annulus throughout the cardiac cycle. This would provide a plane with an approximate short-axis orientation, which had a variable position and orientation over time. This second scheme would provide a more uniform and comprehensive measurement of the velocity profile.

There are a number of anatomical features of the ventricle that have not been modelled for this study but are likely to play a critical role in the development of blood flow patterns. The most important of these are the mitral valve leaflets. These highly dynamic structures directly control the flow of blood and are therefore likely to make a large difference to the flow simulations. Another significant improvement would be the incorporation of the left atrium and its inflow tracts. This would enable the inflow boundaries to be moved away from the ventricular cavity. As such, the flow of blood within the ventricle would become less sensitive to inaccuracies introduced by the boundary conditions imposed. Finally, the internal structure of the ventricle is significantly more complicated than the simplified geometry used in this study. The papillary muscles and the trabeculations form an intricate and dynamic set of structures that both obstruct and promote the flow of blood. It is necessary to investigate the detail with which these features must be modelled such that the simulated flow fields are not significantly affected.

Acknowledgement

This work was sponsored by the British Heart Foundation (QL, FS/2001002).

References

1. Gordon DG. The physics of left ventricular ejection and its implications for muscle mechanics. European J cardiology. 1976; 4(suppl): 87-95.
2. Panerai RB, Smaill BH, Borst C et al, A model of left ventricular function in the denervated heart. J Biomed Engng 1979; 1: 161-71.
3. Chahboune B, Crolet JM. Numerical simulation of the blood-wall interaction in the human left ventricle. European Physical J Applied Physics. 1998; 2: 291-97.
4. Peskin CS, McQueen DM. Cardiac fluid-dynamics, Crit Rev Biomed Eng 1992; 20: 451-59.
5. Redaelli A, Montevecchi FM. Intraventricular pressure drop and aortic blood acceleration as indices of cardic inotropy: a comparison with the first derivative of aortic pressure based on computer fluid dynamics. Medical Engineering & Physics. 1998; 20: 231-41.
6. Gonzalez E, Schoephoerster RT. A simulation of three-dimensional systolic flow dynamics in a spherical ventricle: effects of abnormal wall motion. Annals of Biomed Eng 1996; 24: 48-57.
7. Lemmon J.D. , Yoganathan A.P., Three-Dimensional Computational Model of Left Heart Diastolic Function with Fluid-Structure Interaction. J. Biomed. Eng. 122: p. 109-117.
8. Verdonck P, Vierendeels J, Riemslagh K, Dick E. Left-ventricular pressure gradients: a computer-model simulation. Medical \& Biological Engineering \& Computing, 1999; 37: 511-15.
9. Thomas JD, Weyman AE. Numerical modelling of ventricular filling. Annals of Biomed Eng 1992; 20:19-39
10. Taylor TW, Yamaguchi T. Realistic three-dimensional left ventricular ejection determined from computational fluid dynamics. Med End Phys 1995; 17: 602-08.
11. Jones T.N., Metaxas D.N. “Patient-specific analysis of left ventricular blood flow,” Medical imge computing and computer-assisted intervention - MICCAI’98, 1998; pp.156-166
12. Saber N.R., Gosman A.D., Wood N.B., Kilner P.J., Charrier C., Firmin D.N., “Computational flow modelling of the left ventricle based on in vivo MRI data -- initial experience,” Annals Biomed Eng, 2001; 29:275-283.
13. Long Q, Merrifield R, Yang GZ, Kilner P, Firmin D, Xu XY (2003) “The influence of inflow boundary conditions to the intra left ventricular flow predictions” Journal of Biomechanical Engineering, ASME, 125(6):922-927.

Dynamic Heart Modeling Based on a Hybrid 3D Segmentation Approach

Lixu Gu

Computer Science, Shanghai Jiaotong University, Shanghai, China
gu-lx@cs.sjtu.edu.cn

Abstract. A hybrid 3D segmentation approach is proposed in this paper to perform a physical beating heart modeling from dynamic CT images. A Morphological Recursive Erosion operation is firstly employed to reduce the connectivity between the heart and its neighborhood; then an improved Fast Marching method is introduced to greatly accelerate the initial propagation of a surface front from the user defined seed structure to a surface close to the desired heart boundary; a Morphological Reconstruction method then operates on this surface to achieve an initial segmentation result; and finally Morphological Recursive Dilation is employed to recover any structure lost in the first stage of the algorithm. Every one of 10 heart volumes in a heart beating cycle is segmented individually and finally aligned together to produce a physical beating heart model. This approach is tested on 5 dynamic cardiac groups, totally 50 CT heart images, to demonstrate the robustness of this technique. The algorithm is also validated against expert identified results. These measurements revealed that the algorithm achieved a mean *similarity index* of 0.956. The execution time for this algorithm extracting the cardiac surface from a dynamic CT image, when run on a 2.0 GHz P4 based PC running Windows XP, was 36 seconds.

1 Introduction

Cardiovascular disease becomes one of the leading causes of death for both men and women in the worldwide. However, characterization of myocardial deformation during the systolic contraction is a fundamental step toward understanding the physiology of normal heart and the effects of cardiovascular disease. This effort can lead to more accurate patient diagnosis and potentially reduce the morbidity. An accurate physical beating heart model, which represents all the features of the heart deformation, is considered as a fundamental procedure for cardiac deformation analysis and diagnosis.

Fast and accurate segmentation of a heart volume is a basic operation to perform a physical dynamic cardiac modeling. There are several segmentation algorithms described in the literature to facilitate cardiac image visualization and manipulation [1]-[5]. Most of the researchers paid their attentions to the left and/or right Ventricles from cardiac MRJ images [2]-[5]. Some of them performed in 2D manner and the computing time were rarely mentioned. However, the deformation inspection of the whole heart volume including myocardium is also very important for cardiovascular disease study and endoscopic or minimal access cardiac surgical simulation.

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 237-244, 2004.

A new 3D hybrid segmentation algorithm is proposed here to segment and model a complete beating heart from dynamic cardiac CT images, which operates in a multistage manner to perform segmentation rapidly and precisely. Both the computing time and accuracy of the proposed approach are measured here. This study is improved from our previous researches [6,7].

The rest of the paper is organized as follows: in section 2, we present a brief review and improvement of fast marching method and morphological reconstruction techniques, and propose our multistage hybrid segmentation algorithm. We demonstrate this algorithm and present a validation experiment in section 3. The robustness and accuracy of our approach are discussed in section 4.

2 Multistage Hybrid Segmentation Approach

2.1 Level Set and Fast Marching

The level set method [8] is an interface propagation algorithm. Instead of tracing the interface itself, the level set method builds the original curves (so-called *front*) into a level set surface ϕ (a hyper surface), where the front propagates with a speed F in its normal direction. To avoid complex contours, the current front $\phi(x,y,t=i)$ is always set at zero height $\phi=0$. Hence, the level set evolution equation for the moving hyper surface can be presented as a Hamilton-Jacobi equation:

$$\phi_t + F \mid \nabla \phi \mid = 0 \tag{1}$$

The fast marching method [8] is a special case of the Level Set approach. Suppose we now restrict ourselves to the particular case of a front propagating with a speed F, which is either always positive or always negative. This restriction allows us to simplify the level set formulation. If we assume $T(x,y)$ be the time at which the curve crosses the point (x,y), as shown in Fig.1, the surface $T(x,y)$ satisfies an Eikonal equation where the gradient of surface ∇T is inversely proportional to the speed of the front F:

$$\mid \nabla T \mid F = 1 \tag{2}$$

The fast marching method is designed for problems in which the speed function never changes sign, so that the front is always moving forward or backward and the front crosses each pixel point only once. This restriction makes the fast marching approach much more rapid than the more general level set method.

With respect to rapidly computing a segmentation result, we employ the fast marching method in our approach to perform the initial propagation of a contour from an user-defined seed to an approximate boundary. However, the traditional fast matching method is hard to control overflow when the front propagates near to the contour boundary in many cases. An improved speed term is introduced into our approach, which is based on a global average image gradient instead of many local

gradients. This global speed term can efficiently stop the entire front when most part of the front tends to stable. We applied it to the front speed function at (2):

$$F(x,y,z,t) = F(x,y,z).e^{-\lambda \frac{1}{N} \sum_{(x,y,z)\in I} |\nabla G_\sigma * I(x,y,z)|}, \lambda > 0 \quad (3)$$

where $G_\sigma * I$ denotes the convolution of the image with a Gaussian smoothing filter with standard deviation σ. ∇ and N stands for gradient operation and total number of points in the front, respectively. λ is a positive constant.

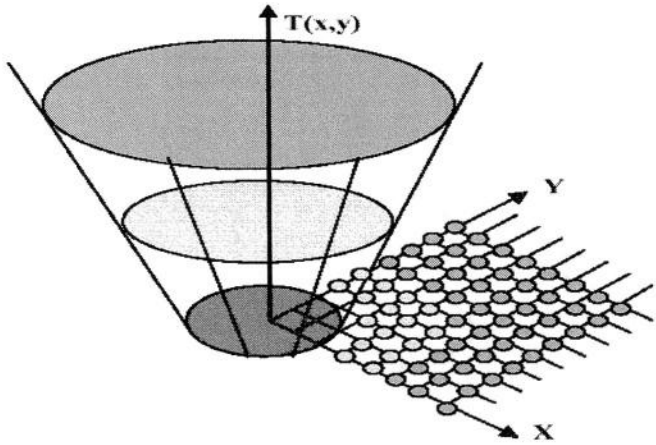


Fig.1 Fast Marching method. T(x,y) demonstrate the time at which the curve crosses the point (x,y).

2.2 Morphological Reconstruction

Mathematical morphology is a powerful methodology for the quantitative analysis of geometrical structures. We employ recursive erosion, dilation and morphological grayscale reconstruction techniques in this research. They are defined below:

Recursive Dilation:

$$F \overset{i}{\oplus} K = \begin{cases} F & if\ i = 0 \\ (F \overset{i-1}{\oplus} K) \oplus K & if\ i \geq 1 \end{cases} \quad (4)$$

Recursive Erosion:

$$F \overset{i}{\ominus} K = \begin{cases} F & if\ i = 0 \\ (F \overset{i-1}{\ominus} K) \ominus K & if\ i \geq 1 \end{cases} \quad (5)$$

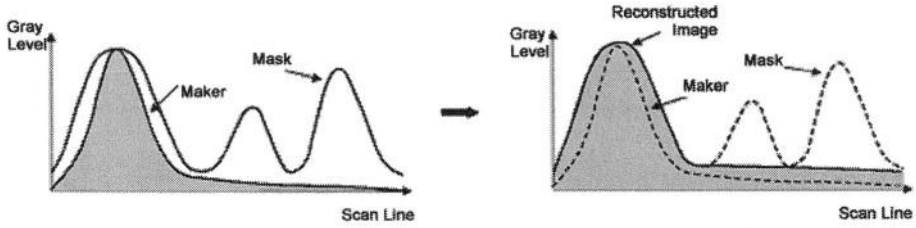


Fig.2 Morphological Reconstruction in grayscale where regions in marker image are used to select regions of the mask image to be reconstructed.

Morphological Reconstruction:

$$B_i = (B_{i-1} \oplus_g k) \bigcap | f |_G \quad (B_i \in R^3, i = 1,2,...) \tag{6}$$

In the above, i is a scale factor and K is the basic structuring element (e.g. 1 pixel radius disk). $\oplus_{gray}$ denotes a dilation operation in grayscale, and $| f |_G$, represents the *mask* of the operation, achieved via a threshold operation using a gray level G. The iteration in (6) is repeated until there is no further change between B_{i-1} and B_i.

Recursive Erosion is employed here to reduce connectivity of objects from neighboring tissues while Recursive Dilation recovers the region lost during the reduction after the objects have been totally segmented. Each step employs the same number of iterations N.

Morphological Reconstruction is a very accurate and efficient tool for recovering the object on a pixel-by-pixel basis. The seed, which results from the output of the fast marching algorithm, is recursively grown under the supervision of the mask until it converges to a stable shape. Morphological reconstruction operations on a grayscale image are depicted in Fig.2.

2.3 Segmentation and Modeling Approach

The proposed 3D segmentation and modeling algorithm is a multistage procedure, comprising 5 major stages:

Stage 1. *Reduce the connectivity between the object region and the neighboring tissues.* Recursively erode the input 3D image using a structuring element base (e.g. a sphere with 1 pixel radius) until the desired object region is completely separated from the neighboring tissues. Record the iteration number i for later use in stage 3. This stage is designed to prevent overflow during the propagation in stages 2 and 3.

Stage 2. *Perform initial evolution of the front.* The improved fast marching method is employed here to initially propagate the user-defined seed to a position close to the boundary without overflow. It performs rapidly, typically less than 10 seconds for a 256×256×100 volume, running on a 2.0 GHz P4 based PC.

Stage 3. *Refine the contours created in stage 2.* Since the speed function in the fast matching method falls to zero sharply, the front could stop a few voxels away from the real boundary. Here, a gray scale morphological reconstruction algorithm is

employed to refine the front as a "final check". The output from stage 2 is employed as a *marker,* while the original image is used for the *mask.*

Stage 4. *Recover the lost data elements from stage 1.* During the recursive erosion in stage 1, part of the object (usually around the edges) is also often eliminated. To recover these lost components, the recursive dilation method is employed. The reconstructed object surface is dilated recursively using the same number of iterations i as recorded in stage 1, which results in the recovery of the object surface to the "original" position, ensuring a highly accurate result.

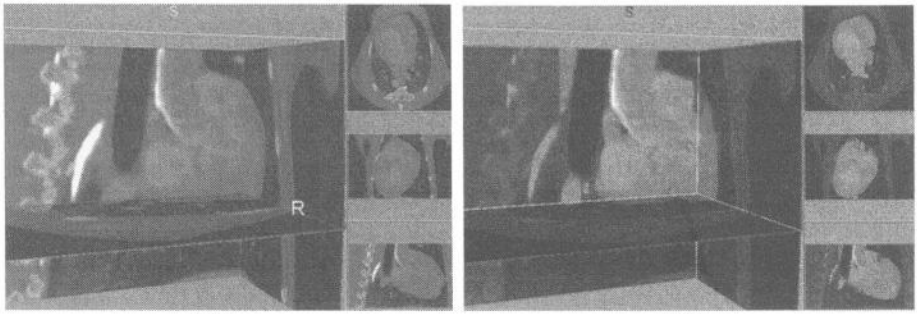
Stage 5. *Model the beating heart.* Finally, a series of dynamic CT volumes involving in a cardiac cycle (totally 10 sets in our experiments) are segmented individually, and the resultant heart volumes are visualized and animated by either surface or volume-rendering methods.

3 Experimental Results

A segmentation environment, "TkSegment", is developed based on the Visualization Toolkit (VTK) and the Python language, into which the multistage hybrid segmentation and modeling algorithm was integrated.

The source data employed in our experiments include 50 CT datasets from heart studies. Five groups of canine CT datasets were employed for the cardiac modeling study. Each was a dynamic volume, acquired with a gated acquisition technique on an 8-slice GE helical CT scanner, consisting of 86 slices at each of 10 equally spaced snapshots during the cardiac cycle. The images were each 512×512 pixels (0.35mm×0.35mm), with an axial spacing of 1.25 mm. One example of them is shown in Fig.3 (a).

The proposed segmentation and modeling algorithm was applied to these 50 volume datasets. A 2.0 GHz P4 based PC running MS-windows XP was employed to run the segmentation. The results of these experiments are described below.



(a) (b)

Fig.3 An example of the cardiac images. (a) source data; (b) highlighted segmented heart region. In both (a) and (b), left: ortho-planar view; right top to bottom: axial, sagittal and coronal views.

3.1 Case Study

Five dynamic CT scans of beating hearts, each containing ten individual volumes throughout the cardiac cycle, were employed in this study. Each of the 10 images was segmented individually. An example of the segmented results is shown in Fig.3 (b).

The average segmentation time for one of these volumes is 155 seconds, which is not fast enough due to the additional time required to segment the blood vessels. However, if the blood vessels are removed early in pre-processing, computational time reduces dramatically to 36 seconds.

The segmented heart volumes were visualized using a Ray Cast volume rendering method and animated to produce a physical cardiac beating model. One set of the heart volumes during a cardiac cycle is shown in Fig.4.

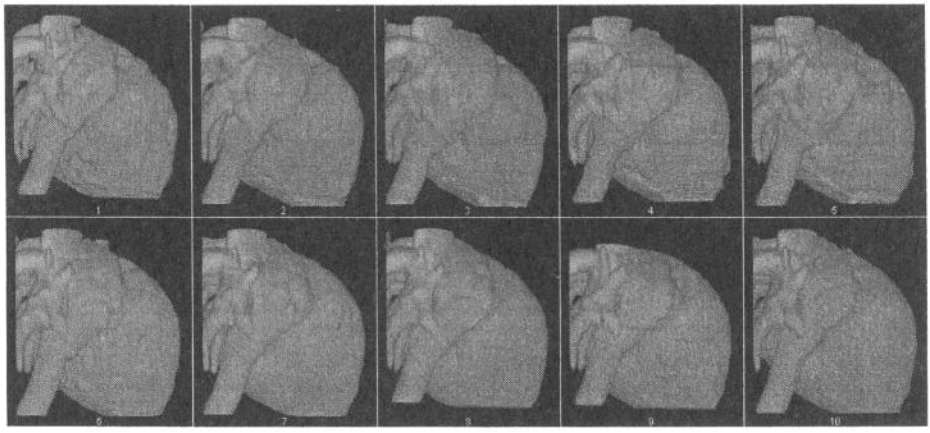


Fig. 4 Segmentation results. 1-10: segmented heart volumes during a cardiac cycle.

3.2 Validation

The segmentation results on the 5 experimental datasets were examined by eye, and deemed to be sufficiently accurate.

To quantify the segmented results, we used the *similarity index* definition introduced by Zijdenbos[9], where manually traced contours were employed as the gold standard. An average *similarity index* of 0.956 was obtained from the heart segmentation study.

4 Discussion

This approach achieves highly accurate segmentation results for the input datasets. Our method identifies and reconstructs the structure of the organ for high quality visualization across a variety of conditions, even in the imperfect world of routine clinical-quality images. Additionally, we believe that it represents the first near real time, full 3D segmentation approach.

Robustness of the multistage hybrid segmentation approach was tested by 5 cardiac datasets in dynamic CT modality. No failed segmentations were reported even in low quality clinical images. Based on our test using VTK build-in algorithms and running in the same computing environment, morphological operations alone require 9 minutes to segment a individual heart volume. Compared to existing segmentation algorithms, our new approach represents a significant improvement.

The hybrid approach has been optimized for 3D volume segmentation. It combines the speed advantage of model-based methods with the high accuracy of region-based methods, resulting in an algorithm that is both fast and accurate. Over all our experiments, segmentations achieve a mean similarity index of 0.956.

The physical beating heart models were finally produced using the segmented cardiac volumes. The animated beating heart can represent the features of deformation of the heart during a cardiac cycle.

5 Conclusion

A new fully 3D medical image segmentation and modeling approach was described using a fast multistage hybrid algorithm. The algorithm takes advantage of the speed and accuracy of both model-based and region-based segmentation methods. It was tested on 5 dynamic cardiac CT datasets, demonstrating excellent segmentation results. Quantitative validation demonstrated an average similarity index of 0.956.

While the procedure currently requires a minimal user-interaction to place seeds, we propose to improve the algorithm to make it fully automatic. We are considering the morphological Top-hat transformation [10], which can extract regions of high intensity of similar size to the objects to be segmented. The detected regions can then be employed as initial seeds. However, this step is still quite computationally expensive, and we therefore chose not to use it in our current work.

Acknowledgement

The author would like to thank Professor T.Peters for his enthusiastic advice and support. Thank P. Lang and J.Xu for their help in validation of the segmentation approaches. We are also grateful to Drs. T.Lee, and D. Gobbi for their help with this project. This work was supported in part by the Ontario Consortium for Image-Guided Therapy and Surgery, and the Canadian Institutes for Health Research.

References

1. A. Frangi, W. Niessen, and M.A. Viergever. "Three-Dimensional Modeling for Functional Analysis of Cardiac Images: A Review". *IEEE Trans. Med. Imag.,* 20(1): pp.2–25, 2001.
2. M.A.Guttman, J.L.Prince and E.R.McVeigh, "Tag and Contour Detection in Tagged MR Images of The Left Ventricle", IEEE Trans. Med. Imag., 13(1): pp.74–88, March 1994.

3. M.R.Kaus, J.V. Berg, W. Niessen and V.Pekar, "Automated Segmentation of the Left Ventricle in Cardiac MRI", MICCAI 2003, LNCS 2878, pp. 432–439, 2003.
4. S.C. Mitchell, B.P.F. Lelieveldt, R.J. van der Geest, H.G. Bosch, J.H.C. Reiber, and M. Sonka. "Multistage hybrid active appearance model matching: Segmentation of left and right ventriclesin cardiac MR images". IEEE Trans. Med. Imag., 20(5): pp.415–423, 2001.
5. K.Suzuki, I.Horiba, N.Sugie and M.Nanki, "Extraction of Left Ventricular Contours From LeftVentriculograms by Means of a Neural Edge Detector", *IEEE Trans. Med. Imag.,* 23(3): pp.330–339, March 2004.
6. L.Gu and T.Peters: An Accurate and Efficient Hybrid Approach for Near Real Time 3D Brain Segmentation, Proc. Of 17th International Congress and Exhibition On Computer Assisted Radiology and Surgery, London UK, pp.1309, June. 2003.
7. M.A. Audette, T.M. Peters, "Level-set segmentation and registration for computing intrasurgical deformations", Proc. SPIE 3661 Medical Imaging, pp.110-121,1999.
8. J.A. Sethian., "Level Set Methods and Fast Martching Methods", Cambridge University Press, 1999.
9. A. P. Zijdenbos, B. M. Dawant, R. A. Margolin, A. C. Palmer, "Morphometric Analysis of White Matter Lesions in MR Images: Method and Validation", IEEE Trans. Med. Imag., vol.13, pp.716-724, Dec. 1994.
10. L.Gu, T.Kaneko, "Extraction of Organs Using Three-Dimensional Mathematical Morphology", Systems and Computers in Japan, Vol.31-7, pp.29-37, 2000.

Tag Stripes Tracking from Cardiac MRI by Bayesian Theory

Min Tang[1], Yuan-Quan Wang[1], Pheng Ann Heng[2], and De-Shen Xia[1]

[1] Department of Computer Science & Technology, Nanjing University of Science and Technology, Nanjing 210094
lilytang@vip.sina.com

[2] Department of Computer Science & Engineering, The Chinese University of Hong Kong, Hong Kong, Shatin

Abstract. Tag tracking is a pre-step to heart motion reconstruction. In this paper, we present a new tag tracking method based on Bayesian statistical approach. our method works on the basis of tracking with the active grid model, it builds the Markov Random Field (MRF) model according to the prediction of the position of the grid node, and classifies the nodes into two categories considering whether they are in the left ventricle using the EM algorithm. then, different prior distribution and likelihood function are designed for different sorts. The iterated conditional modes (ICM) are utilized to maximize the posterior estimate. The method was validated on several sequences of cardiac systole MRI images. Experiment shows that the method can accurately track the SPAMM tag lines without manually outlining the myocardium, and the grid model can keep its topology during tracking process.

Keywords: Markov Random Field, tag stripes tracking, Bayesian theory.

1 Introduction

Tagged magnetic resonance image is an important non-invasive method in the community of cardiac motion analysis in recent years. In this paper, the cardiac MRI images tagged by SPAMM (Spatial Modulation Magnetization) pattern are used. SPAMM can bring out a group of spatial tag planes at cardiac end-diastole. The tag planes are perpendicular to image planes, and the black cross lines of those planes are tag lines (see Fig.2 (a)). Because the tag lines move along with the motion of the tissue, their deformation can represent the tissue motion in image plane.

Using active contour model (Snake) to track tag is popular method in recent years [2-4]. This kind of method divides the tags into two stacks according to directions, and tracks the tag lines one by one in each stack. We think the method based on Snakes suffers from the following two drawbacks: first, it is difficult to exactly set the parameters of every energy terms; in [3], Kumar set different inter-energy parameters at tag intersections to prevent the snake bend, second, most cardiac motion reconstruction algorithms only need the motion information of tag intersections, but

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 245-252, 2004.

methods using Snake can't get those points' motion information directly. Using grid model to track the SPAMM mesh can regard the grid as the prior shape, and directly track the tag intersections according to the grid nodes, so the methods using grid model outperform those using Snake.

Amini used the coupled B-Snake grids to track tag [5]. The algorithm designed the energy of the B-Snake grid intersections and image energy of B-spline, and tracked tag by minimizing the total energy. But the algorithm encountered two difficulties in implementation: first, the algorithm needs to outline the endo- and epi-cardium in each frame by hand; second, the grid model doesn't take into account the connection between nodes, which makes the model fail to punish excessive deformation and maintain the topological shape.

We propose a new tracking method based on Bayesian probability theory which calculates the new grid nodes' coordination to track the tag grids by maximizing the Bayesian posterior probability. We classify the nodes into two categories due to their position in the ventricle or tissue. In this method, the position of each node in the second frame is forecasted, then we classify them using MRF model and EM algorithm. We design different energy functions for each category of nodes according to their function in tracking process, which can make the nodes in the ventricle moves along the nodes in the tissue, while not affecting the nodes in tissue to track the tag intersections. We also take the MRF property of the grid model into account, so our method can retain the topological shape during tracking process.

2 MRF Grid Model to Track Tag

Geman [6] gave MRF image model and utilized it in image restoration in 1984. The nodes in grid model only correlate with its adjacent eight nodes and connected lines, so the grid model bears the MRF property that a node only correlates with its neighborhood. We use the MRF model to estimate these nodes' coordinate to track the tag grid. $S = \{s_i \mid i = 1,...,n\}$ is the node set of the grid model, $\mathrm{L} = \{l_{i,j} \mid i \in S, j \in S, i < j\}$ is the line set, $l_{i,j}$ is the line connecting nodes s_i and s_j. $Q = \{q_i \mid i = 1,..., n\}$ is the coordinate set, q_i is the coordinate of the node s_i. $\alpha = \{\alpha_i \mid i = 1,..., n\}$ is the sort of the node. $L = \{l \mid l = 1, -1\}$ $\alpha_i \in L$, $\alpha_i = -1$ denotes node s_i is in ventricle, otherwise $\alpha_i = 1$. We can get the function $P(Q|Y) \propto P(Y|Q)P(Q)$ based on Bayesian theory, Y is an image observation. The nodes' coordinate can be calculated by MAP: $\hat{Q} = \arg\max_{Q \subset \Omega} P(Q \mid Y)$.

Because of the Markov property that the prior distribution of a site only correlates with its neighborhood, and Hammersley-Clifford theorem that a MRF can equivalently be characterized by a Gibbs distribution, the prior distribution can characterized by a Gibbs distribution as follows:

$$P(Q) \propto \exp(-U_p(Q)),\ U_p(Q) = \frac{1}{Z}\sum_{c \in C} V_c(q_i, i \in c)$$

where Z is a normalizing constant called the partition function, U_p is a prior energy function, and c is a clique. In this paper, we only consider two-point clique and two-order neighborhood of the grid model (see fig.1). We divide the cliques into two sorts c_L and c_{NL}, according to whether there is a grid line between two nodes of the clique, and design different clique potential energy function V_c for them. The potential energy function is also different for nodes in the tissue and ventricles.

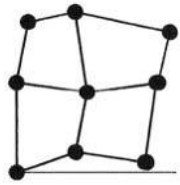


Fig. 1. Two-order neighborhood of grid model

2.1 Clique Potential Energy for Nodes in Tissue

According to the image protocol, the ratio of the tag length to the tag width is small, and the deformation of the tag is within $\pm 10°$. When the adjacent nodes in tissue exactly track the tag intersections, we can consider the line between the two nodes is exactly on the tag. Because the grayscale of the tag is low, and the model shape should be maintained during the tracking process, we design the clique potential energy of node in tissue as below:

$$V_{C_L} = \frac{\left(D\left(l_{i,j}\right) - D\left(l_{i,j}^0\right)\right)^2}{2\sigma_L^2} + mean\left(I\left(l_{i,j}\right)\right),\quad V_{c_{NL}}(i,j) = \frac{\left(D\left(l_{i,j}\right) - D\left(l_{i,j}^0\right)\right)^2}{2{\sigma_L}^2}$$

where $D(l_{i,j})$ is the length of line $l_{i,j}$, $mean\left(I(l_{i,j})\right)$ is the mean intensity of the line $l_{i,j}$, $l_{i,j}^0$ means the undeformed line between nodes s_i and s_j, and σ_L is the deviation of the line length which set the region where the length can change.

2.2 Clique Potential Energy for Nodes in Ventricle

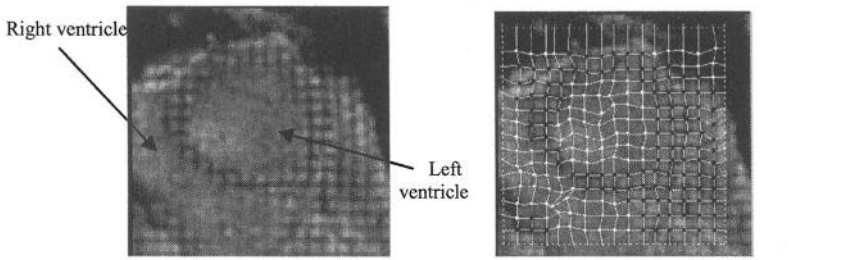
The shape of the ventricle particularly changes during cardiac deformation. If completely ignoring the energy of the nodes in ventricle, the nodes will not change their positions by minimizing energy, which will also influence the motion of the nodes in tissue. Based on above analysis, we consider the nodes in the ventricle should satisfy two requirements, first, they can change positions follow the motion of nodes in tissue, second, their motion can't change the grid model's topological shape.

Smoothness assumption of the optical flow field viz., the optical flow changes smoothly in most of the regions in image plane. Utilizing this assumption and MRF

model, we can make the nodes in ventricle have the approximately similar motion-vector (optical flow) as the adjacent nodes. We design their potential energy as:

$$V_c(i,j) = \frac{\left(D\left(l_{i,j}\right) - D\left(l_{i,j}^0\right)\right)^2}{2{\sigma_L}^2} + \frac{\left|\mathbf{v}_i - \mathbf{v}_j\right|^2}{2{\sigma_\mathbf{v}}^2}, \text{ and } \mathbf{v}_i = \overrightarrow{q_i^0 q_i}, \quad c = \{c_L, c_{NL}\}$$
$$+ \left(\frac{\alpha_j + 1}{2}\right) mean\left(I(l_{i,j})\right)$$

where V_i, V_j are the motion vectors in nodes s_i and s_j respectively, q_i^0 is the initial position of the node s_i in the frame, $\sigma_\mathbf{v}$ is the deviation of the motion vector. When the adjacent point is the node in tissue, the mean intensity of the line between them should be calculated, because the partial tag line between them also should be tracked exactly.



(a) The second frame of the first image plane. (b) Classification result and forecasted grid of the (a).

Fig. 2. Result of forecasted grid and classifying nodes

2.3 Likelihood Function of the Model

Likelihood function $P(Y \mid Q)$ represents the probability of observation Y when given the coordination Q of the grid model. Only tracking the tag intersections, we define the likelihood function as the probability that positions of intersections are the coordinates of the grid nodes.

Fisher [7] and Kraitchman [8] all utilize cross-correlation algorithm to detect tag intersections. We use the profile template of tag line [9], and spread the profile in two imaging direction to get the tag intersection template f'. We can set the likelihood function as:

$$P(Y \mid Q) \propto \exp(-U_L), U_L = \sum_{i=1}^{n}\left(1 - \frac{(\alpha_i + 1)}{2} \cdot \frac{\sum_{k=1}^{N} f'_k N_k(q_i)}{\|f'\|\|N_k(q_i)\|}\right)$$

where U_L is the likelihood energy, the term behind the sign "$\cdot$" is the simplify cross-correlation function, $(\alpha_i + 1)/2$ shows that the likelihood energy only be calculated in

tissue-nodes, N is the size of the intersection template. A correlation result can be seen in Fig.2. To track the tag intersections, we need to look for the coordinate of grid nodes to minimize the total energy.

3 Classify Nodes of Grid Model

In order to automatically track the tag grid without needing outlining the contour of myocardium, we classify the nodes into two sorts due to in ventricle or tissue. The tag is added to the heart at end-diastole (the first frame), so the tag is almost undeformed in the first frame. The blood flow in the ventricle fetches the tag pattern away, so the tag in ventricle will disappear in the followed frames (see fig.2 (a)), while the tag intersections in tissue will not disappear. We can classify the nodes in the second frame: first, forecasting the exact position of the nodes, then, classifying the nodes by utilizing the image feature and the MRF model. The result of classification will be used in the followed frames without changed.

Referencing Zhang's algorithm to segment the brain MRI image [10], and only considering the image feature in the grid nodes, we design the classifying method. We can estimate the sort by MAP: $\hat{\alpha} = \arg\max_{\alpha \in \Omega} P(Y \mid \alpha) P(\alpha)$.

1) Likelihood Function

Taking the two independent observations of image Y_1 and Y_2 into account, we can get the equation: $P(Y \mid \alpha) = P(Y_1 \mid \alpha) P(Y_2 \mid \alpha)$.

The Gaussian likelihood distribution:

$$P(Y_j \mid \alpha) = \prod_{i=1}^{n} f(y_{j,i}, \theta_j^{\alpha_i}), \quad f(y_{j,i}, \theta_j^{\alpha_i}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\left(\frac{y_{j,i} - \mu_j^{\alpha_i}}{\sqrt{2}\sigma_j^{\alpha_i}} \right)^2 - \log\left(\sigma_j^{\alpha_i}\right) \right)$$

where j = 1, 2 represents the sort of the observations, i = 1,..., n is the subscript of the nodes, and if the Gaussian emission function is assumed for the observable random variable Y_j , the mean and standard deviation of each Gaussian class are the parameters, so that $\theta_j^l = \left(\mu_j^l, \sigma_j^l\right)$.

2) Prior Distribution

MRF model can take the spatial information into segmentation algorithm. Considering the two-order neighborhood of grid model (see fig.1), we design the prior distribution as:

$$P(\alpha) = \prod_{i=1}^{N} P\left(l \mid \alpha_{N_i}\right) = \exp\left(-U'(\alpha)\right), \quad U'(\alpha) = \frac{1}{Z} \cdot \beta \sum_{i=1}^{n} \sum_{j \in N_i} (-\alpha_i \alpha_j)$$

where N_i is the two-order neighborhood of node s_i.

3) Parameter Estimation

Because the parameter θ is unknown, we use the EM algorithm to estimate the parameter. Directly combining the E step (Calculating expectation) and the M step

(Maximizing the expectation to get new parameter estimation), we can gain the iterative equation of the parameter:

$$\left(\mu_j^l\right)^{t+1} = \frac{\sum_{i=1}^{n} P^t\left(l \mid y_{1,i}, y_{2,i}\right) y_{j,i}}{\sum_{i=1}^{n} P^t\left(l \mid y_{1,i}, y_{2,i}\right)}, \quad \left(\sigma_j^l\right)^{t+1} = \sqrt{\frac{\sum_{i=1}^{n} P^t\left(l \mid y_{1,i}, y_{2,i}\right)\left(y_{j,i} - \mu_j^l\right)^2}{\sum_{i=1}^{n} P^t\left(l \mid y_{1,i}, y_{2,i}\right)}}$$

where $P^t\left(l \mid y_{1,i}, y_{2,i}\right) = \frac{\prod_{j=1,2} f^t\left(y_{j,i}, \theta_j^t\right) P^t\left(l \mid \alpha_{N_i}\right)}{P(y_{1,i}, y_{2,i})}$.

By analyzing the image feature in the tag intersections, we define the observable random variables $y_{1,i}$ $y_{2,i}$ as:

$$y_{1,i} = \frac{\sum_{k=1}^{N} f'_k N_k(q_i)}{\|f'\| \|N_k(q_i)\|}, \quad y_{2,i} = D\left(N_k(q_i)\right)$$

where $y_{1,i}$ is cross-correlation of the tag template with the region that the center is the node s_i and the size is N (see the Chapter 2.3). $y_{2,i}$ is the intensity deviation of the same region. Fig.2 (b) is the classification result of Fig.2 (a), in which the nodes in tissue are labeled in black, and the others (nodes in ventricle or lung) are white. We can see all the nodes are classified exactly, except for four black nodes in the left bottom of the figure.

4 Experiments and Conclusion

For each image plane, we only apply our method to temporal images acquired during systole, and use the first frame of every slice to form the initialized grid model. Because of the tag pattern added to the heart at end-diastole, the tag lines in first frame are almost undeformed, and vertical to each other in two tag directions. We can project the intensity in two tag directions in the ROI (region of interest) of the image, and approximately locate the position of the tag lines by looking for the valley of the projection, so we can automatically initialize the grid model in the first frame.

4.1 Result of Experiment

The heart images used in the experiment are acquired with the Siemens 1.5 T clinical MR system. The image acquisition matrix dimension is 280×256 with a pixel size of $1.4 \times 1.4 mm^2$ while the slice thickness is 6mm. The reference frame is taken at end-diastole(ED), and there are 16 time-frames in each cardiac cycle. Because the tag grid in diastole process is too blurry to track, we only track tag during systole period. The whole task only needs giving region of interest, then, it can be automatically performed. In Fig.3, we give the result of applying the Amini's algorithm and ours to the same frame, which indicates that our method can restrict the distances of adjacent

nodes and exactly track the nodes without outlining myocardium. For Amini uses the four-order B-spline grid, if one node's coordinate is changed, nearly the total energy needs to be recalculated, and our method only needs to calculate the changes of the energy of one node, so the time complexity of our method is lower. Tracking one frame by Amini's and our method take about 1220s and 100s respectively(MATLAB 6.5, PIII 1G, 256M Cache).

We estimate our algorithm by comparing our result with manually given intersections' positions. We calculate the distances between corresponding nodes, and use the mean distance to evaluate the tracking error. In Fig.4, we present the error(s) calculated from all the SA data.

4.2 Conclusion

We design a new method to track the tag in cardiac MRI image based Bayesian theory. The method takes the spatial connection of grid nodes into account, so the model can keep its shape during tracking process; we design different Bayesian prior probability and likelihood function, which can make the nodes in ventricle or cage move with those in tissue; the method classifies the grid nodes automatically, so we don't need to manually segment the myocardium.

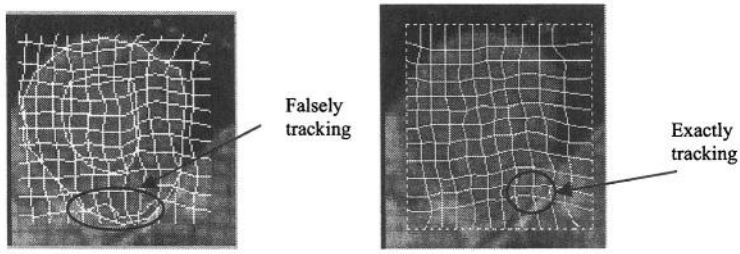


Fig. 3. Result of tracking tag in the last frame of the second slice.

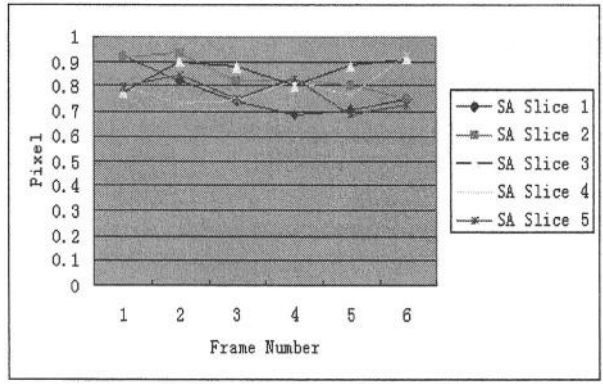


Fig. 4. SA tracking result compared with manually labeled tag intersections.

References

1. L. Axel, *et al.* MR imaging of motion with spatial modulation of magnetization. Radiology, 1989, 171(3): 841-845
2. A. A. Young. Tracking and finite element analysis of stripe deformation in magnetic resonance tagging. IEEE Transactions on Medical Imaging, 1995, 14(3): 413-412
3. Senthil Kumar and Dmitry Goldgof. Automatic tracking of SPAMM grid and the estimation of deformation parameters from cardiac MR images. IEEE Transactions on Medical Imaging, 1994, 13(1): 122-132
4. Thomas S. Denny. Estimation and detection of myocardial tags in MR image without User-Defined myocardial contours. IEEE Transactions on Medical Imaging, 1999, 18(4):330-344
5. Amir. A. Amini, Yasheng Chen and Rupert W. Curwen, *et al.* Coupled B-Snake grids and constrained Thin-Plate splines for analysis of 2-D tissue deformations from tagged MRI. IEEE Transactions on Medical Imaging, 1998, 17(3): 344-356
6. Stuart Geman, D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1984, 6(6):721-741
7. D. Fisher. Automated tracking of cardiac wall motion using magnetic resonance marker. Iowa City: University of Iowa, 1990
8. Dara. L. Kraitchman, A. A. Young, *et al*.Semi-Automatic tracking of myocardial motion in MR tagged images. IEEE Transactions on Medical Imaging, 1995, 14(3): 422-433
9. William S. Kerwin, Nael F. Osman and J. L. Prince. Image processing and analysis in tagged cardiac MRI. In: I. Bankman ed. Handbook of Medical Imaging, New York: Academic Press, 2000, Chapter 24
10. Yongyue Zhang, Michael Brady and Stephen Smith, Segmentation of brain MR images trough a hidden Markov random field model and the Expectation-Maximization algorithm. IEEE Transactions on Medical Imaging, 2001, 20(1):45-57

Determination of the Intracranial Volume: A Registration Approach

Sven Hentschel and Frithjof Kruggel

Interdisziplinäres Zentrum für Klinische Forschung (IZKF),
Inselstrasse 22, D-04103 Leipzig, Germany
kruggel@cbs.mpg.de

Abstract. An algorithm to segment the intracranial compartment from PD-weighted MR images of the human head is described. If only a T_1-weighted dataset is available for a given subject, an artifical PD-weighted dataset is computed from a dual-weighted reference by non-linear registration of the T_1-weighted datasets, and the intracranial compartment is segmented from this artificial dataset. The performance of the algorithm is evaluated on the basis of 12 dual-weighted datasets with an average volume difference of 2.05% and an average overlap (Dice index) of 0.968.

1 Introduction

Skull growth occurs along the suture lines and is determined by brain expansion, which takes place during the normal growth of the brain [9], [19]. Thus in normal adults, a close relationship between the brain size and the intracranial volume (ICV) is expected. This relationship is used to estimate the premorbid brain size in degenerative brain diseases (e.g., Alzheimer's disease [7], [11], [21]) or brain degeneration due to diffuse or focal brain damage.

Three major approaches were suggested to determine the ICV from images of the head: (a) manual delineation of the intracranial compartment in CT [1], [10], [18] or MR images [7], [8], (b) classification and segmentation of multispectral MR images [2], [3], [4], and (c) classification and segmentation of T_1-weighted MR images [15], [16]. While a manual delineation is certainly laborious, the second approach requires the acquisition of multispectral volume images, which is often too time consuming (and thus, too costly) to be acceptable for routine clinical studies. The third method, using T_1-weighted images only, makes certain assumptions that are invalid at least for datasets acquired by our imaging protocol.

Our approach is based on the idea that proton-density (PD)-weighted MR images provide a good basis for ICV segmentation, because the skull signal intensity is low, and all intracranial tissue and the cerebrospinal fluid (CSF) provide a high signal intensity. Thus, the first part of our algorithm consists of generating an ICV mask from a PD-weighted MR image. Most often, only a high-resolution T_1-weighted MR image is available. So the second part of our algorithm consists of a non-linear registration of a T_1-weighted reference image to a T_1-weighted study image, yielding a field of inter-subject deformation vectors. This deformation field is applied to the PD-weighted

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 253–260, 2004.

reference image to generate an "artificial" PD-weighted study image. This artificial PD-weighted image is finally segmented to yield an ICV mask for the study image.

In the next section, we describe our approach in more detail. Then, we evaluate its performance in a "bootstrap" fashion. Finally, we compare our method and results with the three approaches mentioned above.

2 Algorithms

In the following, certain heuristics (detailed below) require that the dataset has been aligned with the stereotactic coordinate system. The x axis corresponds to the ear-to-ear direction, the y axis to the nose-to-back direction, the z axis to the head-to-feet direction. Indices *ca* resp. *cp* refer to the position of the anterior and posterior commissure.

2.1 Generating an ICV Mask from a PD-weighted MR Image

On input, we expect a PD-weighted image of a human head at an isotropical resolution of 1 mm and an intensity resolution of 256 steps. Intensity inhomogeneities should have been corrected by any suitable algorithm. Note that the dataset has been aligned with the stereotactic coordinate system, e.g., by registration with an aligned T_1-weighted image of the same subject. The algorithm consists of three steps: (a) computation of a head mask, (b) computation of an initial ICV mask, (c) refinement of the ICV mask at the brainstem and hypophysis.

Computation of a head mask: The aim is to segment the head region as a single connected component without holes. Steps are spelled out as follows:

$i_1 = \text{isodata}(i_{PD}, 2)$ // segment into two intensity classes
$i_2 = \text{binarize}(i_1, 1, 1)$ // select foreground voxels
$i_3 = \text{dilate}(i_2, 5)$ // morphological dilation by 5mm
$i_4 = \text{invert}(i_3)$ // next 3 steps fill holes inside head mask
$i_5 = \text{selectBig}(\text{label}(i_4, 26))$ // select biggest 26-connected component
$i_6 = \text{invert}(i_5)$
$i_7 = \text{erode}(i_6, 5)$ // restore original head size
$i_{hm} = \text{selectBig}(\text{label}(i_7, 26))$ // select the biggest component

Computation of an initial ICV mask: The next step is to generate a first mask of the intracranial region. The threshold *th* and eroding distance *dist* are determined iteratively:

$i_1 = \text{erode}(i_{hm}, 5)$ // erode head mask by 5mm
$i_{ext} = \text{invert}(i_1)$ // this mask contains all exterior voxels
$th = 40, dist = 4$ // set initial parameters
do {
  $i_2 = \text{binarize}(i_{PD}, th, 255)$ // select voxels above an intensity threshold
  $i_3 = \text{erode}(i_2, dist)$ // separate ICV from small components
  $i_4 = \text{selectBig}(\text{label}(i_3, 26))$ // select biggest 26-connected component
  $i_{icv1} = \text{dilate}(i_4, dist)$ // restore original mask size
  $th \mathrel{+}= 5, dist \mathrel{+}= 0.6$ // increment parameters
} while $(\text{and}(i_{icv1}, i_{ext}) \neq \{\})$ // stop if exterior and ICV mask do not overlap

Due to the application of morphological operators with large kernels, some areas of the intracranial volume with high curvature are rounded off in this first mask. A refinement step adds these voxels back:

$i_1 = \text{dilate}(i_{icv1}, 2)$ // dilate ICV mask by 2mm
$i_2 = \text{mask}(i_1, i_{PD})$ // mask out these voxels from the PD image
$i_3 = \text{binarize}(i_2, th, 255)$ // select voxels above an intensity threshold
$i_4 = \text{and}(\text{invert}(i_{ext}), i_1)$ // select only voxels above an intensity threshold...
$i_5 = \text{open}(\text{and}(i_3, i_4), 1)$ // ...that do not belong to the exterior mask
$i_{icv2} = \text{selectBig}(\text{label}(i_5, 26))$ // select biggest 26-connected component

Computation of the final ICV mask: The brainstem and hypophysis regions need special treatment, because here the high flow in large vessels and CSF lead to a low signal in the PD-weighted image. Thus, parts in these areas are not included in the initial ICV mask.

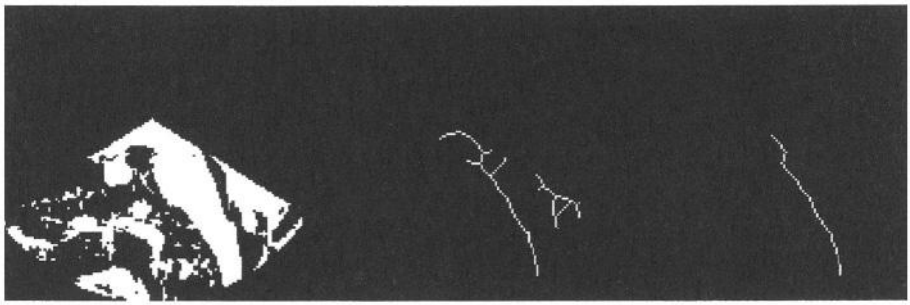


Fig. 1. Refinement of brainstem segmentation: Midsagittal plane of image i_{cone} (left), after thinning (middle), and brainstem midline (right).

Refinement at the brainstem: A cone-shaped mask is placed with a tip at the center of the anterior and posterior commissure, and a basis at the bottommost slice of the dataset with a maximum radius of 80mm. Voxels of the PD-weighted image within this mask are selected if their intensity is above 100 to yield the binary image i_{cone} (see Fig. 1). The medial surfaces of the objects in this image are computed [20]. Sagittal slices in this datasets are searched for the longest connected line that is denoted as the brainstem midline.

Using this line, the brainstem and its surrounding CSF is segmented sequentially in axial slices. In each slice, the smallest distance d_h from the midline voxel v_m to the next background voxel is determined. All foreground voxels within the circle of radius d_h around v_m are collected as the initial brainstem mask i_{bs}. This mask is dilated by 10mm to yield i_{bs_10}. In this mask i_{bs_10}, each foreground voxel is visited radially, starting from the midline voxel. A foreground voxel is eliminated if one of the following conditions is true: (a) this voxel belongs to the background in i_{PD}, (b) we approached the dura mater around the brainstem: this voxel does not belong to the brainstem in i_{bs}

and has an intensity above 80 in i_{PD}, (c) condition (a) or (b) were already true on the radial path.

Finally, a morphological opening using a 1.5mm kernel and a selection of the biggest connected component leads to the brainstem mask that is joined with i_{icv2} to yield i_{icv3}.

Refinement at the hypophysis: In the aligned images, the position of the hypophysis is well known within the subvolume ($x_{ca} - 20 \leq x \leq x_{ca} + 20, y_{ca} - 20 \leq y \leq y_{cp}$, $z_{ca} \leq y \leq 160$). Voxels within this subvolume above the threshold *th* of i_{PD} are collected as image i_{hy}. Now, the region around the hypophysis is segmented as follows:

$i_1 = \text{and}(\text{invert}(i_{icv3}), i_{hy})$ // remove voxels that already belong to i_{icv}
$i_2 = \text{open}(i_1, 1)$ // remove small bridges to the hypophysis
$i_3 = \text{selectBig}(\text{label}(i_2, 26))$ // select the hypophysis
$i_4 = \text{or}(i_3, i_{icv})$ // join the hypophysis with i_{icv}
$i_{icv4} = \text{close}(i_4, 5)$ // close small gaps

Finally, starting from the bottommost axial slice in image i_{icv4} upwards, the area of the ICV mask is calculated. A running average is computed over 10 slices; the level z_{cer} at which the current value is greater than 2 times the running average is taken as the basis of the cerebellum [15]. Voxels in axial slices below $z_{cer} + 10$ are removed to yield the final ICV mask i_{icv} (see Fig. 2).

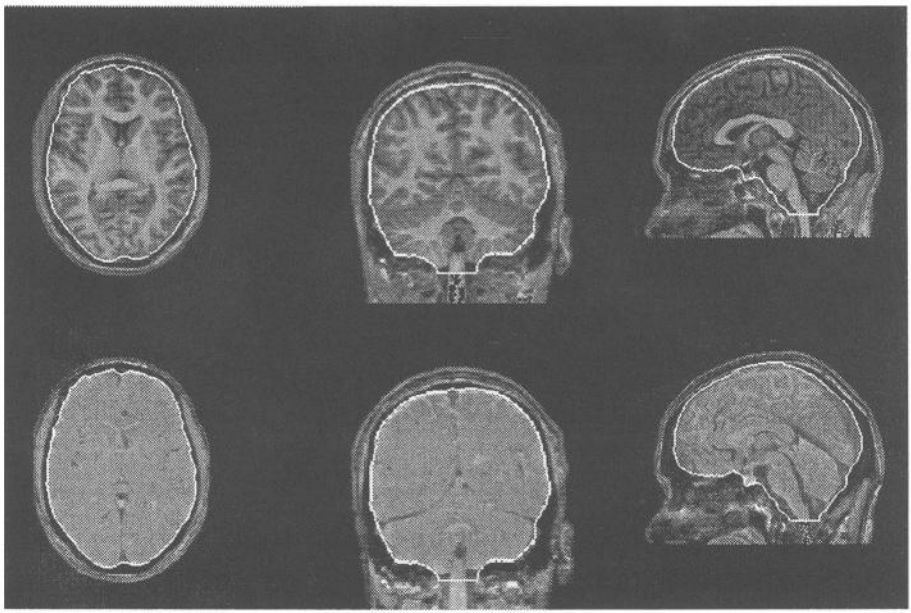


Fig. 2. Border of the ICV mask overlaid on the T_1-weighted image (top row) resp. PD-weighted image (bottom row) for reference sample 12 (see below).

This algorithm was implemented in C++ using the BRIAN environment [12]. The computation time is 221s (AMD Athlon 1800+ machine, Linux 2.4 operating system).

2.2 Generating an Artificial PD-weighted Image

Given reference datasets i_{T1_ref} and i_{PD_ref}, an artificial PD-weighted image i_{PD_stu} for a study subject is computed using i_{T1_stu}. A non-linear registration from i_{T1_ref} onto i_{T1_stu} yields a field of deformation vectors i_{def}. In principle, any method for non-linear registration may be used here that accomodates large-scale deformations. We used an approach based on fluid dynamics [5], [23]. The deformation field i_{def} is applied to the reference dataset i_{PD_ref} to yield an artificial PD-weighted image i_{PD_stu}. Our registration algorithm was implemented in C++ using the BRIAN environment, the computation time is about 22min (due to image dependent optimization, measured on an AMD Athlon 1800+ machine, Linux 2.4 operating system).

In summary, if T_1- and PD-weighted datasets are available for the same subject, an ICV mask is generated using the first algorithm. If only a T_1-weighted dataset is available, an artifical PD-weighted dataset is computed from a dual-weighted reference by non-linear registration, and an ICV mask is segmented from this artificial dataset.

3 Evaluation

Subjects: The MPI maintains a database of subjects enrolled for functional MRI experiments. Before admission, a brief history and physical inspection is taken by a physician. Subjects are included in this database if they comply with the informed consent for conducting general fMRI experiments, pass the examination and do not exhibit pathological features (e.g., unilateral ventricular enlargements, subarachnoidal cysts) in their MR tomograms. Twelve subjects were selected, for which high-resolution T_1- and PD-weighted datasets were available, generally acquired in separate sessions.
Image Acquisition: Magnetic resonance imaging (MRI) was performed on a Bruker 3T Medspec 100 system, equipped with a bird cage quadrature coil. T_1-weighted images were acquired using a 3D MDEFT protocol [14]: FOV 220×220×192 mm, matrix 256×256,128 sagittal slices, voxel size 0.9×0.9 mm, 1.5 mm slice thickness, scanning time 15 min. PD-weighted images were acquired using a 3D FLASH protocol with the same resolution parameters.
Preprocessing: T_1-weighted images were aligned with the stereotactical coordinate system [13] and interpolated to an isotropical voxel size of 1 mm using a fourth-order b-spline method. Data were corrected for intensity inhomogeneities by a fuzzy segmentation approach using 3 classes [17]. PD-weighted images were registered with the aligned T_1-weighted images (6-parameter transformation for rotation and translation, normalized mutual information cost function, simplex optimization algorithm). Finally, the registered PD-weighted images were corrected for intensity inhomogeneities using 2 classes.
Processing: Data of one subject were considered as a reference. Artificial PD-weighted images were computed for the other 11 subjects by the method described above. ICV

	1	2	3	4	5	6	7	8	9	10	11	12
ΔV	1.68	1.82	2.03	2.11	3.54	2.25	1.93	1.73	2.14	1.46	2.73	1.23
dc	0.969	0.970	0.966	0.967	0.962	0.961	0.971	0.964	0.968	0.972	0.969	0.971

Table 1. Averaged volume differences ΔV (in percent) and overlap *dc* (Dice similarity index) for each reference vs. the 11 study subjects.

masks were determined from the real and the artifical PD-weighted images. Their volume differences ΔV (in percent) and overlap *dc* (as measured by the Dice similarity index [6]) were computed. So in total, 12 by 11 comparisons were made. Note that a low volume difference (< 2%) and a high Dice index (> 0.96) correspond to a good adaptation of the ICV mask. Averaged results for each reference are compiled in Table 1. The volume difference ranged between 0.02% and 8.69%, the Dice index between 0.934 and 0.981. Best results were achieved if using sets 10 or 12 as reference.

Results Discussion: Although the algorithm may appear complex at first sight, it requires a set of only 10 basic image processing operations. The validity of the built-in anatomical heuristics were carefully checked for our database, and are expected to be valid for any (normal) MR image of the head.

Several factors influence the ICV segmentation: (a) The quality of the reference datasets. Head motion, flow and DC artifacts impede good segmentation results, (b) A high flow in the sinuses may lead to a low signal at the border of the intracranial cavity in the PD-weighted image, leading to possible segmentation errors at the ICV border. However, the induced volume error was found to be less than 0.5%. (c) In areas of the convexity of the skull where the tabula interna is very thin, the partial volume effect may smear the signal intense dura mater with the bone marrow, so that parts of the bone marrow are included in the ICV mask. Again, only a small induced volume error (0.2%) was found. In summary, the ICV mask should be checked visually when selecting datasets as a reference.

Other factors influence the adaptation quality of the artificial ICV mask: (a) We noted a significant relation between the volume difference ΔV before and after registration, e.g. a difference in the ICV volume between the reference and the study of 200ml leads to a volume error ΔV of 40ml (or 3%) in the artificial ICV mask. Most likely, this is a consequence of the partial volume effect in the registration procedure, since the ICV border layer has a volume of typically 65ml. (b) One may ask whether the deformation field generated from a non-linear registration of T_1-weighted datasets is a good model for the anatomical inter-subject differences, and thus suitable for applying it to the PD-weighted dataset. In particular, this is true for study cases where we found a low Dice index. In summary, the ICV difference between reference and study image should be small to yield a good ICV estimate for the study dataset.

In practice, one or more reference datasets should be chosen from a larger group by the method discussed above. Selection criteria are a low volume difference (< 2%) and a high Dice index (> 0.96) for all adaptations in the group. The mean error may be used as an estimate for the expected error in the generation of the study ICV mask.

4 Discussion

A new approach for the determination of the intracranial volume in MRI datasets of the human head was described. In a nutshell, an ICV mask is computed from a high-resolution PD-weighted dataset. If such an image is not available, a non-linear registration between a T_1-weighted dataset of a reference and a study subject yields a deformation field that is applied to the reference PD-weighted dataset in order to obtain an artificial study PD-weighted dataset. An ICV mask for the study subject may then be generated. Using a suitable reference, this approach yields an expected volume error of less than 2% and an overlap of better than 0.97. The process is fully automatical and reliable: On a 4 processor cluster, we generated ICV masks for a database of 540 normal subjects in 68h.

Compared with the three approaches mentioned in the introduction, manual delineation of the intracranial cavity, as previously used in our [21], [22] and other studies [1], [7], [8], [10], [11], [18] is tedious (about 1.5h of work per dataset). If performed by an expert, it may still be considered as the gold standard, although small ambiguities due to the partial volume effect and inter-rater variability induce a volume error of the same magnitude as our method.

Alfano *et al.* [2], [3] suggested to use multispectral MRI datasets for ICV segmentation, while Lemieux *et al.* [15], [16] base their approach on T_1-weighted data only. Our method lies somewhat between both of these approaches: we use the helpful information provided by the PD-weighted datasets for ICV segmentation, but do not require that multispectral data are available for all subjects in a study. If high-resolution T_1-weighted data are provided, this method may even be used retrospectively.

As noted in the introduction, the ICV is closely related to the brain size of young healthy adults. Thus, ICV measures may be used to estimate the premorbid brain size, which is useful to compute the amount of atrophy in brain degeneration due to diffuse diseases (e.g., Alzheimer's disease, anoxic encephalopathy, microangiopathy) or following focal brain damage (e.g., cerebral infarction or hemorrhage, after tumor removal).

Acknowledgement

The authors wish to thank the MPI of Human and Cognitive Brain Science, Leipzig, for providing the datasets.

References

1. Abbott, A.H., Netherway, D.J., Niemann, D.B., Clark, B., Yamamoto, M., Cole, J., Hanieh, A., Moore, M.H., David, D.J.: CT-determined intracranial volume for a normal population. J Craniofac. Surg. **11** (2000) 211–223.
2. Alfano, B., Brunetti, A., Covelli, E.M., Quarantelli, M., Parnico, M.R., Ciarmiello, A., Salvatore, M.: Unsupervised, automated segmentation of the normal brain using a multispectral relaxometric magnetic resonance approach. Magn. Reson. Med. **37** (1997) 84–93.

3. Alfano, B., Quarantelli, M., Brunetti, A., Larobina, M., Covelli, E.M., Tedeschi, E., Salvatore, M.: Reproducibility of intracranial volume measurement by unsupervised multispectral brain segmentation. Magn. Reson. Med. **39** (1998) 497–499.
4. Andersen, A.H., Zhang, Z., Avison, M.J., Gash D.M.: Automated segmentation of multispectral brain MR images. J Neurosci. Meth. **122** (2002) 13–23.
5. Christensen, G.E.: Deformable shape models for anatomy. Thesis (1996) Washington University, St. Louis.
6. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology **26** (1945) 297–302.
7. Edland, S.D., Xu, Y., Plevak, M., O'Brien, P., Tangalos, E.G., Petersen, R.C., Jack, C.R.: Total intracranial volume: normative values and lack of assiociation with Alzheimer's disease. Neurology **59** (2002) 272–274.
8. Eritaia, J., Wood, S.J., Stuart, G.W., Bridle, N., Dudgeon, P., Maruff, P., Velakoulis, D., Pantelis, C.: An optimized method for estimating intracranial volume from magnetic resonance images. Magn. Reson. Med. **44** (2000) 973–977.
9. Falkner, F.: Normal growth and development: current concepts. Postgrad. Med. **62** (1977) 58–63.
10. Hahn, F.J., Chu, W.K., Cheung, J.Y.: CT measurements of cranial growth: normal subjects. Am J Roentgenol. **142** (1984) 1253–1255.
11. Jenkins, R., Fox, N.C, Rossor, A.M., Harvey, R.J., Rossor, M.N.: Intracranial volume and Alzheimer disease. Arch. Neurol. **57** (2000) 220–224.
12. Kruggel, F., Lohmann, G.: BRIAN (Brain Image Analysis) - A toolkit for the analysis of multimodal brain data sets. In: Computer Aided Radiology (CAR'96) 323–328, Elsevier, Amsterdam.
13. Kruggel, F., von Cramon, D.Y.: Alignment of magnetic-resonance brain datasets with the stereotactical coordinate system. Med. Imag. Anal. **3** (1999) 1–11.
14. Lee, J.H., Garwood, M., Menon, R., Adriany, G., Andersen, P., Truwit, C.L., Ugurbil, K.: High contrast and fast three-dimensional magnetic resonance imaging at high fields. Magn. Reson. Med. **34** (1995) 308–312.
15. Lemieux, L., Hagemann, G., Krakow, K., Woermann, F.G.: Fast, accurate, and reproducible automatic segmentation of the brain and in T_1-weighted volume MRI data. Magn. Reson. Med. **42** (1999) 127–135.
16. Lemieux, L., Hammers, A., Mackinnon, T., Liu, R.S.N.: Automatic segmentation of the brain and intracranial cerebrospinal fluid in T_1-weighted volume MRI scans of the head, and its application to serial cerebral and intracranial volumetry. Magn. Reson. Med. **49** (2003) 872–884.
17. Pham, D.L., Prince, J.L.: An adaptive fuzzy segmentation algorithm of magnetic resonance images. IEEE T Med. Imag. **18** (1999) 737–752.
18. Sgouros, S., Hockley, A.D., Goldin, J.H., Wake, M.J., Natarajan, K.: Intracranial volume change in craniosynostosis. J Neurosurg. **91** (1999) 617–625.
19. Sgouros, S., Goldin, J.H., Hockley, A.D., Wake, M.J., Natarajan, K.: Intracranial volume change in childhood. J Neurosurg. **91** (1999) 610–616.
20. Tsao, Y.F., Fu, K.S.: A parallel thinning algorithm for 3D pictures. Comp. Graph. Imag. Proc. **17** (1981) 315–331.
21. Wolf, H., Kruggel, F., Hensel, A., Wahlund, L.O., Arendt, Th., Gertz, H.J.: The relationship between head size and intracranial volume in elderly subjects. Brain Res. **973** (2003) 74–80.
22. Wolf, H., Hensel, A., Kruggel, F., Riedel-Heller, S.G., Arendt, Th., Wahlund, L.O., Gertz, H.J.: Structural correlates of mild cognitive impairment. Neurobiol. Aging **25** (2004), in print.
23. Wollny, G., Kruggel, F.: Computational cost of non-rigid registration algorithms based on fluid dynamics. IEEE T Med. Imag. **21** (2001) 946–952.

Shape and Pixel-Property Based Automatic Affine Registration Between Ultrasound Images of Different Fetal Head

Feng Cen, Yifeng Jiang, Zhijun Zhang, and H.T. Tsui

Electronic Engineering Department
The Chinese University of Hong Kong
Shatin, NT, Hong Kong SAR
{fcen, yfjiang, zjzhang, httsui}@ee.cuhk.edu.hk

Abstract. The difficulties in the automatic registration of the ultrasound images of different fetal heads are mainly caused by the poor image quality, view dependent imaging property and the difference of brain tissues. To overcome these difficulties, a novel Gabor filter based preprocessing and a novel shape and pixel-property based registration method are proposed. The proposed preprocessing can effectively reduce the influence of the speckles on the registration and extract the intensity variation for the shape information. A reference head shape model is generated by fusing a prior skull shape model and the shape information from the reference image. Then, the reference head shape model is integrated into the conventional pixel-property based affine registration framework by a novel shape similarity measure. The optimization procedure is robustly performed by a novel mean-shift based method. Experiments using real data demonstrate the effectiveness of the proposed method.

Keywords: ultrasound image registration, shape similarity, gabor filter, mean shift.

1 Introduction

Ultrasound imaging has become the most important medical imaging tool in obstetric examination. It is considered to be a safe, non-invasive, real-time and cost-effective way to examine the fetus. Imaging and measuring the head of the fetus is a key routine examination to monitor the growth of the fetus. The registration of different fetal heads is very useful for comparing the growth of different fetuses and constructing the normalized model of the fetal head for the diagnosis of fetal head malformation.

However, the registration of ultrasound images is more difficult than that of other medical imaging modalities due to the poor image quality of ultrasound images caused by the speckle noise. The methods of medical image registration is typically divided into two categories: feature based methods [1] [2] and pixel-property based methods. As the automatic extraction of the anatomical

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 261–269, 2004.

structure features is quite difficult in ultrasound images, many researches tend to use the pixel-property based methods for the automatic registration of ultrasound images. For example, Meyer *et al* [3] used the mutual information measure to affine and elastic registration, Shekhar *et al* [4] investigated using the preprocessing by median filter and intensity quantization to improve the robustness of the registration and Gee *et al* [5] proposed to use the constraint of the mechanics of freehand scanning process to reduce the computational load in non-rigid registration.

The Biparietal Diameter (BPD) is the maximum diameter of a transverse section of the fetal skull at the level of the parietal eminences. The BPD plane contains the most valuable information for the obstetric doctor to investigate the fetal head and monitor the growth of the fetus. So, in this paper, we shall focus on the automatic registration between the ultrasound images of different fetal heads in the BPD plane.

Actually, the ultrasound image is view dependent, i.e., the structures closely parallel to the ultrasound beam direction will not show up clearly. So, the parts of a skull in the ultrasound beam direction are often invisible in the ultrasound images. Furthermore, in most situation, the difference of the brain tissue between different fetuses is large. Therefore, the conventional pixel-property based methods will fail in our study.

In this paper, we propose a novel shape and pixel-property based method to register the ultrasound images of the BPD plane between different fetuses. In the proposed method, a prior shape model, obtained by hand measurement of a group of fetal head ultrasound images, is used to represent the prior shape information about the skull in the BPD plane. Then, the prior shape model is updated with the reference image to generate a reference shape model. The benefit of combining of the prior shape model and the shape information in the reference image is the more accurate representation of the skull shape even in the case that the skull structure is partly invisible in the ultrasound image. A novel shape similarity measure is proposed to assess the similarity between the shape model and the ultrasound image. Finally, the registration is performed with a linear combination of the shape similarity measure and conventional pixel-property based similarity measure of correlation ratio (CR) [6]. A robust optimization is performed by a novel mean-shift based method. In addition, a Gabor filter based preprocessing is proposed to reduce the influence of the speckles and extract the intensity variation for shape information.

2 Preprocessing

The speckle noises in ultrasound images are able to be viewed as in an irregular and complex texture pattern[7]. This fact inspires us to employ the Gabor filters for the preprocessing of the ultrasound images to reduce the negative impact of the speckle on the performance of registration. The preprocess is illustrated in Fig. 1 (a). First, a wavelet-like Gabor filter bank is constructed to decompose the ultrasound image in spatial frequency space into multiscale and multiorientation. A 2-D complex Gabor filter represented as a 2-D impulse response is given by[9]

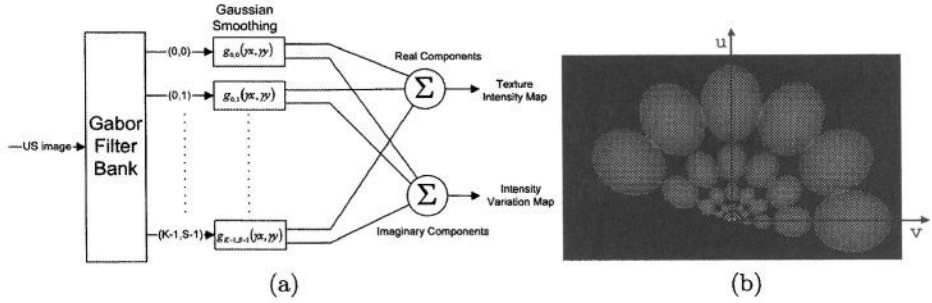


Fig. 1. (a)The preprocessing procedure diagram. (b)The responses of Gabor filter bank in spatial frequency domain. Only the portion larger than the half-peak magnitude is shown for each filter.

$$h(x,y) = \frac{1}{2\pi\sigma_{x'}\sigma_{y'}} \exp\left\{-\frac{1}{2}\left[\frac{x'^2}{\sigma_{x'}^2} + \frac{y'^2}{\sigma_{y'}^2}\right]\right\} \exp(j2\pi F x'), \tag{1}$$

where $(x', y') = (x\cos\theta + y\sin\theta, -x\sin\theta + y\cos\theta)$ are rotated coordinates, F is the radial center frequency and σ_x and σ_y are the space constants of the Gaussian envelope along the x and y axes, respectively.

Let B_F and B_θ denote the frequency bandwidth and the angular bandwidth, respectively. The Gabor filter bank, covering the spatial frequency domain, can be generated by varying four free parameters $(F, \theta, B_F, B_\theta)$.

After Gabor filter decomposition, a Gaussian smoothing is processed for the output amplitude of each channel. The smoothing filter, $g_{k,s}(\gamma x, \gamma y)$, is set to have the same shape as the Gabor filter of the corresponding channel but greater spatial extents. The subscripts $s = (0, ..., S-1)$ and $k = (0, ..., K-1)$ denote the scale and orientation of the outputs, respectively, and the parameter γ controls the spatial extent of the smoothing filter.

In our implementation, we use the parameter set suggested by [8], since the Gabor filters generated with this parameter set have the optimal texture separability. The response of Gabor filter bank in spatial frequency domain is shown in Fig. 1(b).

Finally, compounding the real parts and imaginary parts of the outputs of smoothing filters, respectively, we get

$$G^r(x,y) = \sum_{k,s} H^r_{k,s}(x,y), \qquad G^i(x,y) = \left| \sum_{k,s} H^i_{k,s}(x,y) - \mu_{H^i} \right|, \tag{2}$$

where $H^r_{k,s}(x,y)$ and $H^i_{k,s}(x,y)$ are the real part and imaginary part of the output of $g_{k,s}(\gamma x, \gamma y)$, respectively, and μ_{H^i} is the mean value of $\sum_{k,s} H^i_{k,s}(x,y)$ over the entire image. Since the $G^r(x,y)$ and $G^i(x,y)$ can be considered as the representation of the amplitude of the texture pattern and the variation

of the image intensity, respectively, we call $G^r(x, y)$ the texture intensity map and $G^i(x, y)$ the intensity variation map. To be easily adopted into the pixel similarity measure Eq. 7, the double-valued $G^r(x, y)$ is quantized to 256 levels.

3 Registration

The registration procedure of the proposed method is illustrated in Fig. 2. It consists of two major parts, i.e., the generation of reference shape model and the shape and pixel-property based registration of the ultrasound images.

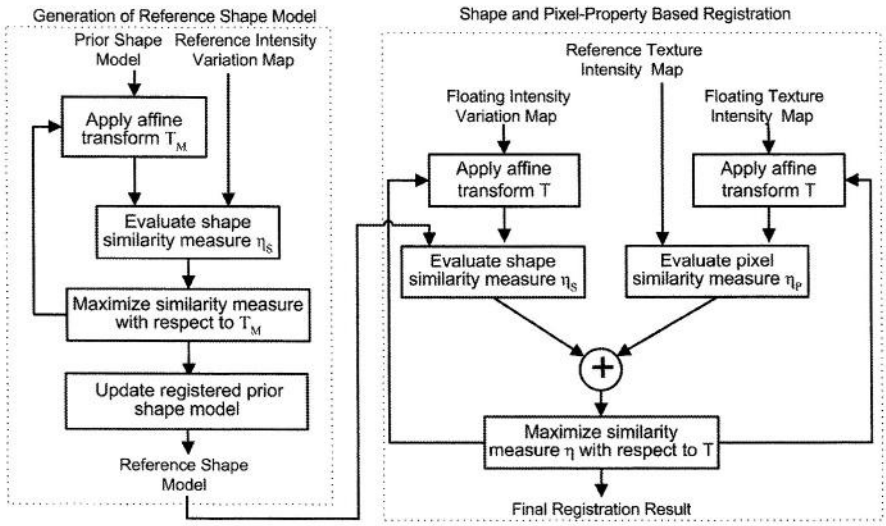


Fig. 2. Diagram of the proposed registration procedure.

3.1 Reference Shape Model Generation

The purpose of this procedure is to build a shape model that can more accurately represent the shape information of the interest organ in the reference image. The shape model, $M(x, y)$, used in this paper is a binary bit map having the same size as the reference image. The regions of value 1 in the shape model is a shape depiction of the object of interest.

The intensity variation map of the reference image and a prior shape model are used to generate the reference shape model. Two steps are involved into this procedure. First, the prior shape model is aligned with the intensity variation map of the reference image by maximize the shape similarity measure with respect to the affine transformation of the prior shape model; then, the shape information extracted from the intensity variation map and the prior shape model are fused to produce the reference shape model.

Here, we propose to use the normalized sum of intensity variation within the region defined by the shape model to assess the similar degree between the object shape in the image and the shape model. The proposed shape similarity measure can be written as

$$\eta_S(M, G^i) = \sum_{x,y} \frac{G^i(x,y)M(x,y)}{G^i_{max}}, \tag{3}$$

where G^i_{max} is the max value of $G^i(x,y)$ over the entire intensity variation map. The range of $\eta_S(M, G^i)$ is from 0 to 1.

Assuming that the prior shape model transformed by a given affine transformation T_M is $M_p^{T_M}(x,y) = M_p(x,y) \circ T_M$, here, $M_p(x,y)$ is the initial prior shape model, the alignment between the prior shape model and the reference image is to seek the T_M^* that maximize the shape similarity measure, i.e.,

$$T_M^* = \arg\max_{T_M}\{\eta_S(M_p^{T_M}, G_r^i)\}, \tag{4}$$

where G_r^i is the intensity variation map of the reference image.

In our study, the prior shape model of the fetal head in BPD plane only takes into account the skull, because it is the most prominent structure in the ultrasound images of the fetal head. The shape of the skull can be modeled as an elliptic strip. The prior shape model, as shown in Fig. 6(a), is acquired by hand measurement of a group of ultrasound images of the BPD plane,

After alignment, the shape information of the reference image is extracted by

$$M_s(x,y) = \begin{cases} 1 & \text{if } G_r^i(x,y) > \alpha\sigma_{H_r^i} \\ 0 & \text{if } G_r^i(x,y) \le \alpha\sigma_{H_r^i} \end{cases}, \tag{5}$$

where $\sigma_{H_r^i}$ is the standard deviation of $\sum_{k,s} H_{k,s}^i(x,y)$ over the entire reference image and α the parameter controlling the extraction of the shape information. In our experiment, the α is set to 1. Then, the reference shape model, an example is shown in Fig. 6 (b), is obtained by

$$M_r(x,y) = M_p^{T_M}(x,y) \oplus M_s(x,y). \tag{6}$$

3.2 Shape and Pixel-Property Based Registration

The similarity measure in the proposed registration method involves two major parts: the shape similarity measure and the pixel-property based similarity measure. The shape similarity measure is the same as Eq. 3. According to our preliminary study, the CR similarity measure have larger extent of the attraction basin than other pixel-property based similarity measure. Moreover, the value of CR measure is comparable to that of the shape similarity measure. Thus, the CR is adopted as the pixel-property based similarity measure in our method.

Suppose that the texture intensity maps of the reference image and floating image are G_r^r and G_f^r, respectively, and the floating image is tansformed by a given affine transformation T. The CR measure is given by [6]

$$\eta_P(G_r^r, G_f^r \circ T) = \frac{Var[E(G_f^r \circ T | G_r^r)]}{Var(G_f^r \circ T)}. \tag{7}$$

Then, the shape similarity measure and pixel-property similarity measure are linearly combined to give the cost function for the registration, i.e.,

$$\eta(I_r, I_f \circ T) = \eta_P(G_r^r, G_f^r \circ T) + \beta \eta_S(M_r, G_f^i \circ T). \tag{8}$$

where β is a weighting constant, I_r the reference image, I_f the floating image.

3.3 Maximization of Similarity Measure

Generally, the desired solution in image registration is related to a strong local maximum of the similarity measure close to the start position, but not necessarily the global maximum. Here, "strong" not only refers to the magnitude of the local maximum, but also means the large extent of the attraction basin. Therefore, we propose to use the Powell's direction set method[10] to provide the direction of each sub-line maximization and the mean-shift algorithm[11] to perform the sub-line maximization. This method can robustly searching for the strong local maximum.

Assuming Ω is a local window around the current location x on the cost function surface with a window radius λ, one dimensional mean-shift based maximizing procedure is iteratively moving the current location by

$$m(x) = \frac{\sum_{s \in \Omega} K\left(\frac{s-x}{\lambda}\right) \eta(s) s}{\sum_{s \in \Omega} K\left(\frac{s-x}{\lambda}\right) \eta(s)} - x, \tag{9}$$

where K is a suitable kernel function.

This iteratively moving procedure is equivalent to hill climbing on the convolution surface given by

$$C(x) = \sum_s H\left(\frac{s-x}{\lambda}\right) \eta(s), \tag{10}$$

where H is the shadow kernel of K. For simplicity and effectiveness, we choose the kernel K as a flat kernel, and the corresponding shadow kernel is the Epanechnikov kernel [11], which is a smoothing kernel and can effectively smooth the local fluctuation on the surface of similarity function. Moreover, by varying the kernel scale λ from large scale to small scale progressively, the optimization can be run robustly and accurately in a coarse-to-fine strategy. An example of the mean-shift maximization in one dimension is shown in Fig. 3.

4 Experiments and Results

Two ultrasound images, as shown in Fig. 5(a), obtained from two different women with around 20-week pregnancy were used in our experiments. The image size is 256 × 256.

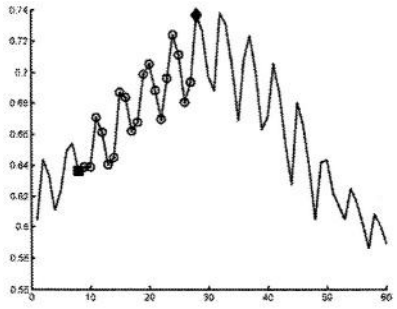


Fig. 3. Searching for maximum with mean-shift method in one dimension. The start position is marked with a solid square, the intermediate steps are shown by hollow circles and the final result is represented by a solid diamond.

The smoothing effect of the Gabor filter based preprocessing on the pixel similarity function is given in Fig. 4. From Fig. 4(b), we can observe that the undesired local maxima in large scale have been effectively smoothed and removed when using the texture intensity maps for the pixel similarity measure. The preprocessing outputs of the reference and floating ultrasound images are shown in Fig. 5 (b) and (c). Note that the texture intensity map after requantization, Fig. 5 (b), can also be view as a despeckled and contrast enhanced ultrasound image.

In Fig. 6, we show the shape model and the final registration result, Fig. 6 (d). In the shape models, the value of white regions is 1. Although there are some small noticeable registration errors, the accuracy of the result is acceptable for the affine registration between two different subjects.

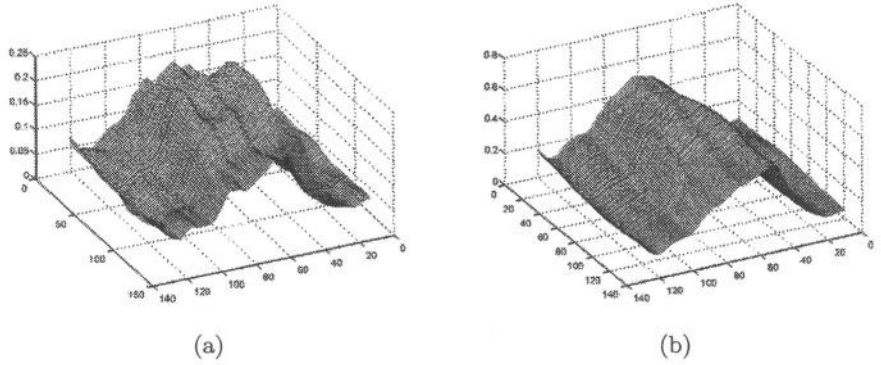


Fig. 4. Correlation Ratio as a function of misalignment caused by translation in y axis and scaling in x axis for (a) the original image pair and (b) the texture intensity maps of the image pair.

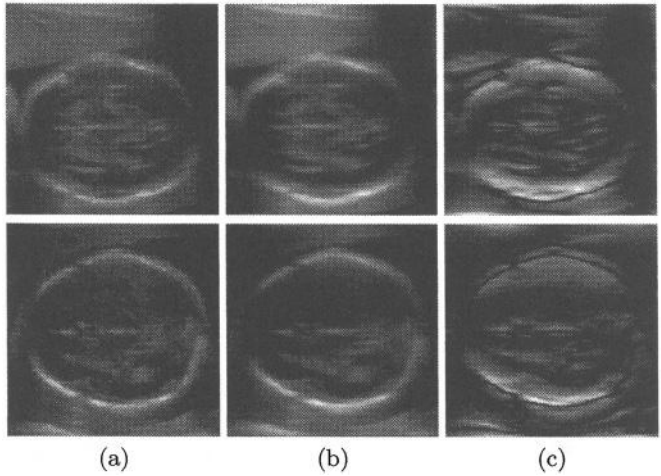


Fig. 5. Results of Gabor filter preprocessing for the reference image (top row) and the floating image (bottom row).(a) Original ultrasound images. (b) Texture intensity maps of the images. (c) Intensity variation maps of the images . For display convenience, it is also quantized to 256 levels.

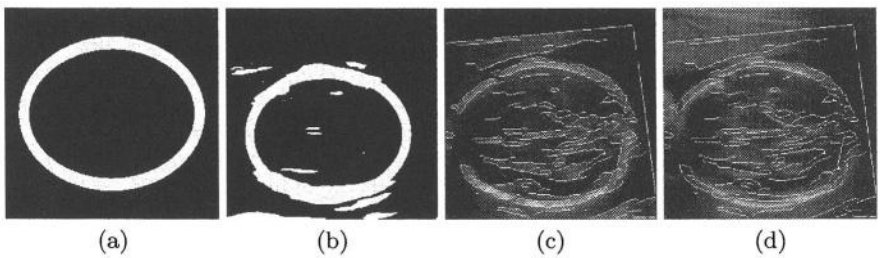


Fig. 6. (a)the prior shape model. (b) the reference shape model. (c)the floating image when registered.(d)the final registration result shown in the reference image. The white curves superimposed on (c) and (d) are the edges detected by Canny edge detector in the corresponding texture intensity map of the floating image.

5 Conclusions

In this paper, we have proposed a novel method for affine registration of the ultrasound images. This method consists of a Gabor filter based preprocessing of the ultrasound images, a shape and pixel-property based similarity measure and a robust searching of strong local maximum based on the mean-shift algorithm. We demonstrate the effectiveness of this method with the experiment

on the registration of the ultrasound images between different fetal heads. The proposed method can be easily extended to the registration of other organs, if the prior shape model of fetal skull is replaced by that of other organs. In our future work, we shall study to apply the proposed method for the registration of other ultrasound images and extend the proposed method to the registration of ultrasound volumes.

References

1. G. Ionescu, S. Lavalle and J. Demongeot. Automated Registration of Ultrasound with CT Images: Application to Computer Assisted Prostate Radiotheapy and Orthopedics. *Proc. MICCAI'99 Springer LNCS,* 1679:768-777,2000.
2. B. C. Porter, D. J. Rubens, J. G. Smith, S. Totterman and K. J. Parker. Three-Dimensional Registration and Fusion of Ultrasound and MRI Using Major Vessels as Fiducial Markers. *IEEE Trans. Medical Imaging,* 20:354-359, 2000.
3. C. R. Meyer, J. L. Boes, B. Kim, P. H. Bland, G. L. Lecarpentier, J. B.Fowlkes, M. A. Roubidoux and P. L. Carson. Semiautomatic Registration of Volumetric Ultrasound Scans. *Ultrasound Med. Biol.,* 25:339-347, 1999.
4. R. Shekhar and V. Zagrodsky. Mutual Imformation-Based Rigid and Nonrigid Registration of Ultrasound Volumes. *IEEE Trans. Medical Imaging,* 21(1):9-22, 2002.
5. A. H. Gee, G. M. Treece, R. W. Prager, C. J. C. Cash and L. Berman. Rapid Registration for Wide Field of View Freehand Tree-Dimensional Ultrasound. *IEEE Trans. Medical Imaging,* 22(11):1344-1357, 2003.
6. A. Roche, G. Malandain, X. Pennec and N. Ayache. The Correlation Ratio as a New Similarity Measure for Multimodal Image Registration. In *Proc. MICCAI'98 Springer LNCS,* 1496:1115-1124, 1998.
7. C. M. Chen, H. H. Lu and K. Han. A Textural Approach Based on Gabor Functions for Texture Edge Detection in Ultrasound Images. *Ultrasound Med. Biol.,* 27(4):515-534, 2001.
8. D. A. Clausi, M. E. Jernigan. Designing Gabor Filters for Optimal Texture Separability. *Pattern Recognition,* 33:1835-1849, 1999.
9. A.C. Bovik, M. Clark and W.S. Geisler. Multichannel Texture Analysis Using Localized Spatial Filters. *IEEE Trans. Pattern Anal. Machine Intell.,* 12(1):55-73, 1990.
10. W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling. Numerical Recipes in C, 2nd ed. *Cambridge, U. K.: Cambridge Univ. Press,* 1992.
11. Y. Cheng. Mean Shift, Mode Seeking, and Clustering. *IEEE Trans. Pattern Anal. Machine Intell.,* 17(8):790-799, 1995.

Multimodal Brain Image Registration Based on Wavelet Transform Using SAD and MI

Jue Wu[1] and Albert C.S. Chung[2]

[1] Bioengineering Program, School of Engineering, the Hong Kong University of Science and Technology, Hong Kong, China
[2] Department of Computer Science, the Hong Kong University of Science and Technology, Hong Kong, China

Abstract. The multiresolution approach is commonly used to speed up the mutual-information (MI) based registration process. Conventionally, a Gaussian pyramid is often used as a multiresolution representation. However, in multi-modal medical image registration, MI-based methods with Gaussian pyramid may suffer from the problem of short capture ranges especially at the lower resolution levels. This paper proposes a novel and straightforward multimodal image registration method based on wavelet representation, in which two matching criteria are used including sum of difference (SAD) for improving the registration robustness and MI for assuring the registration accuracy. Experimental results show that the proposed method obtains a longer capture range than the traditional MI-based Gaussian pyramid method meanwhile maintaining comparable accuracy.

1 Introduction

Image registration is a process of finding a transformation that maps one image onto another same or similar object by optimizing certain criterion. It is an essential step in medical image processing if clinicians require complementary information obtained from different images. Registration aims to fuse data about patients from more than one medical image so that doctors can acquire more comprehensive information related to pathogenesis.

Most of rigid registration methods require an iterative process to go from an initial estimate position to a final optimal one[1]. There are many factors that affect the process of registration. The capture range is one of them; it directly determines the success or failure of the alignment. Regarding intensity similarity measures, many local optima can trap the optimization method and cause it to stop at incorrect positions. Capture range can be defined as a range within which the starting algorithm is more likely to converge to the correct optimum[1]. In other words, if the starting point is not within the capture range, the registration algorithm is most likely to converge to a wrong optimum, thus leading to misregistration. Logically, the size of capture range is positively correlated to the success rate of a registration task. As such, the capture range is a very important property that influences the robustness of registration algorithms and

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 270–277, 2004.

longer capture ranges are what registration methods pursue [2,3]. The size of a capture range depends on the features in the images and similarity metrics used to measure them. In this paper, the capture range is mathematically defined as the range in which the value of matching criterion monotonically decreases (or increases) when at a distance away from the maximal (or minimal) position.

Mutual information (MI) is one of the most popular matching criteria that are used in multi-modal image registration. Many researches have shown that MI produced satisfactory results in terms of accuracy [2,3,4]. Due to its high computational complexity, scientists have proposed the multiresolution scheme to accelerate MI-based registration. Though some researchers believe that a multiresolution scheme can also increase the capture range for there is less tendency to be trapped in local minima [2], our experiments show that the capture range is still not good enough especially in lower resolution registration. This is supported by the conclusion drawn in [5], i.e. the hope that a multiresolution approach to matching would be better equipped to avoid local optima seems unfounded. The statistical relation of image intensities that MI measures tends to decline when the image resolution decreases [6]. Thus, MI does not naturally extend to the multiresolution framework. In order to improve this situation, we propose the combination of the sum of absolute difference (SAD) and MI as similarity metrics rather than MI alone.

We do not use SAD and MI directly on the original image intensities. Instead we make use of the wavelet coefficients to calculate the two similarity measures. Scientists have tried to apply wavelet transform to image registration due to its inherent multiresolution characteristics and ability to preserve significant structural information of images. J.L. Moigne [7] registered remote sensed data using maxima of HL and LH wavelet coefficients. P. Moon *et al.* [8] looked into the applicability of the wavelet transform to digital stereo image matching. Turcajova *et al.* [9] tested various orthogonal and biorthogonal wavelets (they used LL coefficients) together with cross-correlation to register affine transformed images of objects at a distance. Liu *et al.* [10] suggested the application of the Gabor filter and the Gaussian model of registration. J. Zhen *et al.* [11] proposed a wavelet-based multi-scale block matching algorithm to register DSA images.

In this paper, we utilize the LL wavelet coefficients to register two multimodal images using the sum of absolute difference (SAD) at lower resolution levels and mutual information (MI) at higher resolution levels. The originality of this idea is to apply wavelet in the field of medical image registration and the combination of two different similarity measures. Experimental results show that our method has a longer capture range than a conventional Gaussian pyramid with MI at all levels.

2 Proposed Approach

2.1 Matching Criteria

In our algorithm, two similarity metrics are utilized, namely MI and SAD. They work on the LL wavelet coefficients at different resolutions.

Originating from information theory, MI is an entropy-based concept and denotes the amount of information that one variable can offer to the other. In terms of marginal distributions $p(a)$ and $p(b)$ for images A and B respectively and the joint distribution $p(a, b)$, MI can be defined as:

$$I(A, B) = \sum_{a,b} p(a, b) \log \frac{p(a, b)}{p(a)p(b)},$$

where a and b represent the intensity of image A and B respectively. MI measures the statistical dependence between the image intensities of corresponding voxels in both images, which is assumed to be maximal while SAD is assumed to be minimal if the images are geometrically aligned. At a low resolution level, the capture range of MI is not satisfactory and we find that combined with wavelet, SAD works better than MI with a Gaussian pyramid in terms of capture ranges. MI has a shorter capture range because after the MI value drops down from the maximum, it soon increases. This is due to the fact that joint entropy is small when images are registered and it increases considerably (thus MI decreases) when the floating image shifts away from the optimal transformation. When the brain of floating image maps mainly to the background of reference image, the joint entropy decreases. If the amount of decrement is greater than the decrement of the sum of two marginal entropies, MI increases ($I(A, B) = H(A) + H(B) - H(A, B)$). This results in short capture ranges of MI at low resolution levels.

SAD is defined as:

$$SAD = \frac{1}{N} \sum_{\boldsymbol{x} \in D} |u(\boldsymbol{x}) - v(T(\boldsymbol{x}))|,$$

where $\boldsymbol{x}$ and T refer to the position vector and the transformation respectively, u and v are the intensity functions of two images, and N is the number of pixels in the region of interest D (here refers to the overlapping part of the two images).

In a general way, SAD is not effective in multi-modal registration because even when perfectly registered, images from different modalities are different in intensity [2]. Therefore we adopt MI rather than SAD at higher resolution levels to ensure the accuracy of registration. However, SAD works well in lower resolution registration because in a low resolution scenario, all details tend to disappear and only global features are preserved, e.g., a brain image approximates to a ball. An important fact is that in almost all medical images, background is dark and objects of interest are highlighted. This fact excludes the intensity-inverted occasion that is not suitable for SAD to work. This means different medical images from different modalities are inclined to similar intensity profiles, thus SAD can be effective in this scenario. Our experiments show it outperforms MI.

2.2 Choice of Wavelet Coefficients

We use the standard discrete wavelet transform (with Haar basis function which is easy to implement [12]) algorithm to get wavelet decompositions of images

which include four subband images, LL, LH, HL and HH. In our method, we make use of LL subband coefficients to perform registration(See Fig. 1). Rather than the other three subbands which contain details like edges, LL subband preserves most of the significant information one image has, e.g., anatomical structures, intensity values, etc. Thus, as shown in the experiments, it is suitable for MI and SAD to get satisfying results.

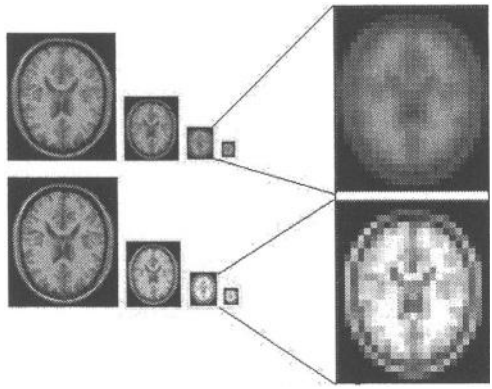


Fig. 1. *Pyramid images. Top left four are Gaussian Pyramid. Bottom left four are wavelet pyramid (LL coefficients). Right are zoomed-in images showing the difference.*

Wavelet transform provides not only frequency domain information but also spatial domain information. A considerable amount of registration work based on wavelet has been done in the remote sensing, cartography, geoscience and so forth. However in the area of medical image registration, little work has involved wavelet transform. We adopt wavelet instead of Gaussian as the representation for two main reasons. Firstly, wavelet transform can keep more significant anatomies in medical images, like the gyrus and ventricles of the brain (See Fig. 1). Secondly, wavelet does not blur the images through the hierarchical pyramid as much as Gaussian filter.

2.3 Multiresolution Registration

We first perform wavelet transform on each image into several levels. The total number of levels of decomposition depends on the image size. Suppose that there are N levels. Level 1 represents the highest resolution level. We start the registration process from the coarsest level, i.e. level N. For every level i $(i > N/2)$, we register the two LL subband images by minimization of SAD and for every level i $(i \leq N/2)$, we register the two LL subband images by maximization of MI. For each i $(N \geq i \geq 2)$, the resulting position of registration of level i is multiplied by 2 and then becomes the initial position of the next level, i.e. level $i-1$. The registration process terminates when the matching criterion is optimized at the highest resolution level.

3 Experimental Results

The testing datasets used in our experiments were obtained from the Brainweb, a Simulated Brain Database (T1, T2 and PD-weighted MR) [13,14], and from the collaborating hospital (MR,CT), where all the corresponding images have been precisely registered (See Fig. 2). We tested the proposed algorithm and MI-based Gaussian pyramid algorithm on 4 groups of multi-modal medical images, T1-T2, T1-PD, T2-PD and MR-CT. In the experiments, we decomposed images into $N = 4$ levels. In the MI-based Gaussian pyramid algorithm, we used MI for the registration of all resolution levels. In our method, we used SAD for the registration of Levels 4 and 3 LL subband images, and MI for the registration of Levels 2 and 1 LL subband images.

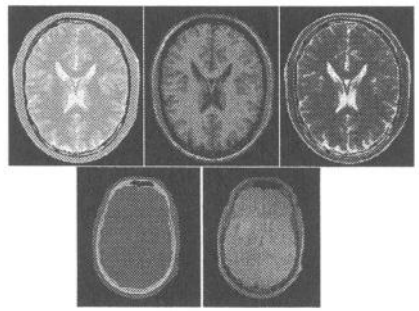


Fig. 2. *Datasets for the experiments. The first column are T1, T2 and PD*(181 × 217 × 181) *images from Brainweb. The second column are MR and CT images*(256 × 256) *from the collaborating hospital.*

In order to study their performances, we plotted the probing graphs of our method and the MI-based method in Figs. 3 and 4 respectively. The number of bins was set to 32. To minimize the interpolation artifact, we adopted the partial volume interpolation (PVI) during the density estimation process [4]. In the figures, capture ranges are delimited by two vertical lines.

Here are some observations about the results. Firstly, in the low resolution translational shifts (Sub-figures (a) and (c) in Fig. 3 and Fig. 4), the new method has a much longer capture range (over double) than the traditional one. This is further justified by the remarkable improvements on different multi-modal image matching groups. The capture ranges along the horizontal direction are given in Table 1. For rotations (Sub-figures (b) and (d) in Fig. 3 and Fig. 4), capture ranges are similar because the overlapping regions of objects do not vary significantly.

Secondly, in high resolution translational shifts (Sub-figures (e) and (g) in Fig. 3 and Fig. 4), the new method has a slightly longer capture range.

Thirdly, in high resolution rotations (Sub-figures (f) and (h) in Fig. 3 and Fig. 4), capture ranges are similar. It is observed that, when the images are aligned,

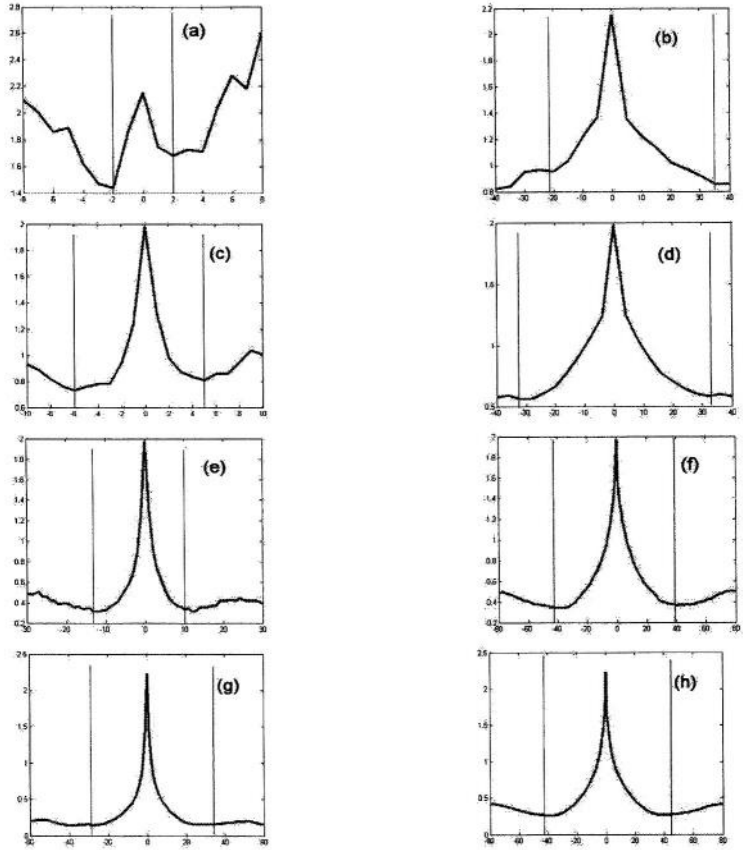


Fig. 3. *The conventional MI-based method with Gaussian pyramid on T1-T2 images. Level 4 is at the top row of the figure while level 1 is at the bottom row. The left column gives the values of MI (All levels) across the translations along the* x *axis and the right column gives the corresponding values for different rotations.*

the MI value of the wavelet LL subband images (around 2.3) is larger than that of the Gaussian filtered images (around 2.1). It shows that the wavelet LL subband images contain more mutual information because the anatomical structures are better preserved in the LL subband images.

4 Summary

In this paper, we present a new multi-modal medical image registration method based on wavelet transform using SAD and MI. Our method extends the short capture ranges with the conventional MI-based Gaussian pyramid method and meantime achieves the same accuracy. The novelty of this idea lies in the combination of two different similarity measures (MI and SAD) in one multiresolution

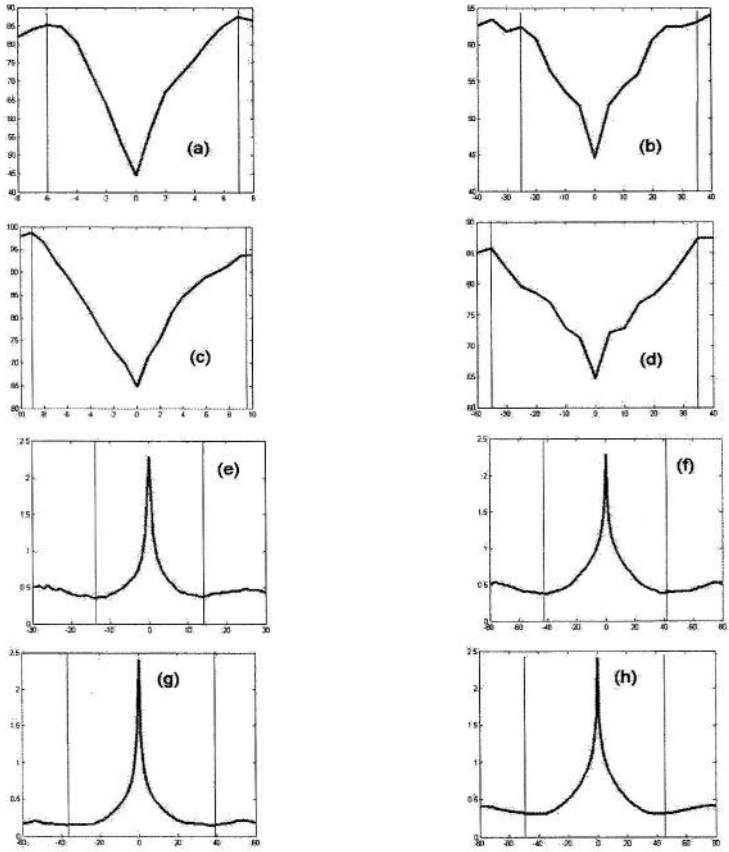


Fig. 4. *Our proposed method on T1-T2 images. Level 4 is at the top row of the figure while level 1 is at the bottom row. The left column gives the values of SAD (Levels 4 and 3) and the values of MI (Levels 2 and 1) across the translations along the* $\boldsymbol{x}$ *axis and the right column gives the corresponding values for different rotations.*

scheme (wavelet pyramid). In principle, this framework can adopt any two metrics and any representation as long as the result is satisfying. Our experiments show that SAD together with wavelet LL subband coefficients obtains a much longer capture range than the conventional method in low resolution registration. Future work will apply wavelet transform to a large number of datasets to further study its performance compared to other existing methods.

References

1. J.Hajnal, D.Hill, D.Hawkes, eds.: Medical Image Registration. 1st edn. CRC Press (2001)

		MI + Gaussian	Our method	Improvements
T1-T2	Level 4	6.7±1.5	13.7±1.1	104%
	Level 3	12±2.4	22.4±3.6	87%
T1-PD	Level 4	8.1±2.9	15.6±2.5	93%
	Level 3	12.9±2.5	26.9±3.3	109%
T2-PD	Level 4	7.1±1.6	14.5±2.2	104%
	Level 3	13.4±2.6	21.7±2.3	62%
MR-CT	Level 4	13.4±2.1	22.1±3.4	65%
	Level 3	30.2±3.9	40.4±7.1	34%

Table 1. *Comparisons of capture ranges in four multimodal image matching groups. For each group, ten slices are randomly selected. Mean and standard deviation are calculated. Probing was performed across translational shifts along the x axis. Unit: pixels.*

2. W.M.III.Wells, P.Viola, H.Atsumi, S.Nakajima, R.Kikinis: Multi-modal volume registration by maximization of mutual information. Medical Image Analysis **1** (1996) 35–51
3. J.P.W.Pluim, J.B.A.Maintz, M.A.Viergever: Mutual-information-based registration of medical images: A survey. IEEE Trans. Medical Imaging **22** (2003) 986–1004
4. F.Maes, A.Collignon, D.Vandermeulen, G.Marchal, P.Suetens: Multimodality image registration by maximization of mutual information. IEEE Trans. Med. Imag. **16** (1997) 187–198
5. J.P.W.Pluim, J.B.A.Maintz, M.A.Viergever: Mutual information matching in multiresolution contexts. Image and Vision Computing **19** (2001) 45–52
6. M.lrani, P.Anandan: Robust multi-sensor image alignment. In: Proc. DARPA Image Understanding Workshop. Volume 1. (1997) 639–647
7. J.L.Moigne, W.J.Campbell, R.F.Cromp: An automated parallel image registration technique based on the correlation of wavelet features. IEEE Trans. Geoscience and Remote Sensing **40** (2002) 1849–1864
8. P.Moon, G.Jager: An investigation into the applicability of the wavelet transform to digital stereo image matching. In: Joint IEEE Communications and Signal Processing Society. (1993) 75–79
9. R.Turcajova, J.Kautsky: A hierarchical multiresolution technique for image registration. In: Proc. of SPIE Mathematical Imaging: Wavelet: Applications in Signal and Image Processing. (1996)
10. J.Liu, B.C.Vemuri, J.L.Marroquin: Local frequency representations for robust multimodal image registration. IEEE Trans. Medical imageing **21** (2002) 462–469
11. J.Zhen, J.Yifeng, Z.Jihong: Automatic registration algorithm based on wavelet transform. In: Proc. of International Conference on Signal Processing 2000. (2000) 979–982
12. R.C.Gonzalez, R.E.Woods: Digital Image Processing. 2nd edn. Prentice Hall (2002)
13. BrainWeb: (Mcgill) http://www.bic.mni.mcgill.ca/brainweb/.
14. C.A.Cocosco, V.Kollokian, R.K.Kwan, A.C.Evans: Brainweb: Online interface to a 3d mri simulated brain database. In: Proceedings of 3rd International Conference on Functional Mapping of the Human Brain. (1997)

Reducing Activation-Related Bias in FMRI Registration

Luis Freire[1,2], Jeff Orchard[3], Mark Jenkinson[1], and Jean-François Mangin[4]

[1] Functional MRI Centre of the Brain, Oxford University, OX3 9DU, Oxford, U.K.
[2] Instituto de Biofísica e Engenharia Biomédica, FCUL, 1749-016 Lisboa, Portugal
[3] School of Computer Science, University of Waterloo, Waterloo N2L 3G1, Canada
[4] Service Hospitalier Frédéric Joliot, CEA, 91401 Orsay, France

Abstract. The presence of cerebral activation may bias motion correction estimates when registering FMRI time series. This problem may be solved through the use of specific registration methods, which incorporate or down-weight cerebral activation confounding signals during registration. In this paper, we evaluate the performance of different registration methods specifically designed to deal with the problem of activation presence. The methods studied here yielded better results than the traditional approaches based on least square metrics, almost totally eliminating the activation-related confounds.

1 Introduction

The problem of subject motion is of particular importance in FMRI studies, due to the small amplitude of the signal changes induced by cerebral activation. Indeed, signal variations introduced by slight movements can easily hide the BOLD signal features related to cognitive tasks, or lead to the appearance of false activations. Due to the lack of perfect and comfortable immobilization schemes, neuroimaging researchers often prefer to rely on retrospective motion correction of the FMRI time series. However, optimal registration of the time-series is not sufficient to correct for all signal changes due to subject motion during data acquisition.

In practice, an important confounding effect may still be introduced during the estimation of motion parameters. This confounding effect relates to the fact that the presence of activation may systematically bias the motion estimates, rendering them correlated with the activation profile (even in the absence of subject motion) [1]. This effect is particularly significant when motion estimates are obtained using similarity measures based on least squares metrics. The consequence may be the appearance of spurious activations along brain edges after statistical analysis. Moreover, this systematic bias is likely to render invalid any attempt to correct other motion-related artifacts by using motion parameters as regressors of no-interest.

A standard approach to reduce the activation-related bias in motion estimates consists of using a robust estimator [2,3], for instance a Geman-McClure

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 278–285, 2004.

M-estimator. In [4], the suitability of this estimator was assess in the context of FMRI registration. This estimator, however, was not sufficient to completely discard activation-related correlations in the estimation of motion parameters. More recently, two new methods dedicated to FMRI registration were proposed in [5,6].

In this paper, we evaluate the robustness of these two least-squares-based registration methods in the presence of activation. The fact that the proposed methods rely on two different computational frameworks, which differ in the interpolation and optimization schemes, has motivated the inclusion of the results obtained by conventional least-squares-based methods implemented under each computational framework. The robustness of the different methods is first evaluated using two artificial time series, produced in order to simulate a situation with absence of subject motion, and another with true activation-correlated motion. The different methods are finally tested on three actual time series obtained from human subjects in a 3T magnet.

2 Materials

2.1 Similarity Measures

In this paper, we have compared four registration methods, which are outlined next:

1. **Least Squares (LS1 and LS2)**. The least squares similarity measure is calculated through the sum of the squared residual difference between voxel pairs, for a given rigid-body transformation, *T*. For two images *A* and *B*, yielding the intensity pair of values (a,b) for voxel i, the least squares cost function is defined as

$$LS(A, B; T) = \sum_i (A_i - B_i^T)^2, \tag{1}$$

in which B_i^T is the resampled value of image *B* at voxel i after the geometric transformation *T* has been applied. The LS registration methods are based on two distinct computational frameworks. The first uses cubic-spline interpolation [7] under a Powell optimization scheme [8]. This LS implementation will be referred to as LS1. The second method relies on the computation of exact intensity derivatives introduced by a simple fixed-point iteration method. This second LS implementation will be referred to as LS2.
To prevent each method from being trapped in local minima, spatial smoothing is applied in both methods by convolving the image with a 3D Gaussian kernel with FWHM value set to 8 mm.
2. **Dedicated Least Squares (DLS)**. This is the registration method that imposes the strictest down-weighting of voxels that potentially bias motion estimates. It relies on a dedicated approach, which runs on a two-stage registration. After the first stage, which is identical to LS1, a rough mask intended

to include all activated areas, is obtained. During the second motion estimation, the voxels inside a dilated version of the activation mask are discarded from the calculation of the similarity measure. It should be noted that this mask may include some spurious activations stemming from a biased initial motion correction, which shall not be a problem if spurious activations are not too wide. An illustration of a discarding mask is provided in Figure 1 *(left)*.

3. **Simultaneous Registration and Activation Detection (SRA).** The fourth method is the SRA, which attempts to overcome the problem of activation bias by incorporating the effects of the activation into the registration problem.
Least-squares registration involves finding the motion parameters x that minimize the sum of squared differences between the volumes A and B. Using a linear approximation to the motion transformation, this means solving

$$A = B + Gx, \tag{2}$$

in the least-squares sense, where the matrix G is the derivative of the transformation with respect to the motion parameters. In this case, consider the volumes A and B to be column vectors. Similarly, the formulation of activation detection using the general linear model,

$$A = B + yH, \tag{3}$$

lends itself to solving for the coefficients of activation, y, in a least-squares sense. Here, A and B can be thought of as voxel time series stored as row-vectors, while H holds the stimulus regressors in its rows. Since the volumes are stored in columns, and voxel time series are stored in rows, an entire dataset can be stored in a single matrix. Using this notation, an FMRI dataset can be modeled as a baseline dataset (B), plus a motion component (GX), plus an activation component (YH). Hence, we have the model:

$$A = B + GX + YH. \tag{4}$$

A least-squares solution (X, Y) can easily be found using matrix factoring techniques or by an iterative method. However, it can be shown that the solution is not unique. Adding a regularization constraint that favours sparse activation maps has been shown to work on simulated FMRI datasets [6].

3 Description of the Experiments

3.1 FMRI Acquisitions

We have used a set of three FMRI studies, corresponding to three different subjects. The images were acquired on a Bruker scanner operating at 3T. Volume geometry is (64 × 80 × 18, 3.75 mm × 3.75 mm × 6.00 mm). A block design was used to assess visual activation, in which two tasks were repeatedly presented to subjects following a box-car design. Each task had a duration of 18 s (corresponding to 9 volumes), and was repeated 10 times, yielding 180 images.

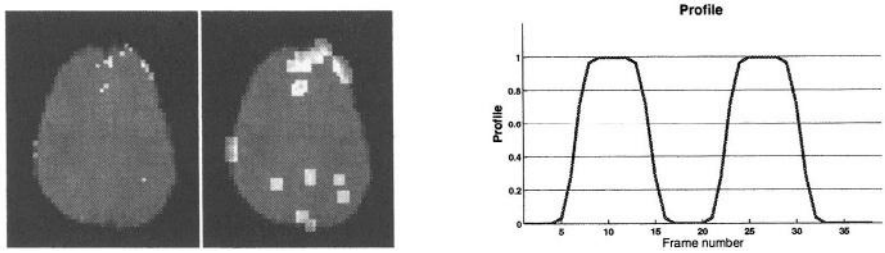


Fig. 1. *(Left)*: example of the DLS discard mask with the estimated activated regions and activation mask obtained by dilation of activated regions. *(Right)*: activation profile used to generate the simulated time series.

3.2 Experiments with Simulated Time Series

Motionless Simulated Time Series: The evaluation of the different registration methods is first performed using an artificial time series designed to simulate an activation process in the absence of subject motion. This was done by duplicating a reference image forty times and adding an activation-like signal change to some voxels in order to mimic a cognitive activation process. The activation time profile is shown in Figure 1 *(right)*.

The added activation pattern was obtained using SPM99, after resampling the first actual time series with LS1 motion estimates (the activation pattern has a size of 6.7% of the brain and the mean activation level was set to 2% of brain mean value). To simulate the effects of thermal noise, Rician noise was added to the dataset by applying Gaussian noise (standard deviation of 2% of the mean brain value) to both the real and imaginary components. The four registration methods are then applied. For each registration method, we also compute the Pearson's correlation coefficient of the 6 estimated motion parameters with the cognitive task profile. For the DLS method, the activation mask was obtained by statistical inference of the resampled simulated time series, according to [5].

The results for the simulated time series are presented in Figure 2. One can see that the DLS and the SRA registration methods can effectively reduce the bias in motion estimates introduced by the presence of activation. Substantial reductions in the correlation coefficients can also be observed in Table 1 *(left)*.

Simulated Time Series with True Correlated Motion: The elaboration of this simulated time series is similar to the previous one, except that true activation-correlated motion was added. This is not an uncommon situation in real studies (for instance, if the subject is asked to speak) and, in fact, the consequences of a poor alignment may be disastrous in this situation. The simulated true correlated motion, which comprises translations in x and y directions, follows the same profile as the activation, with maximum amplitude in both directions of 2 mm. In order to minimize interpolation-related artifacts, Fourier interpolation in k-space was used, which is in accordance with the 2-D ($x - y$ plane) signal acquisition process.

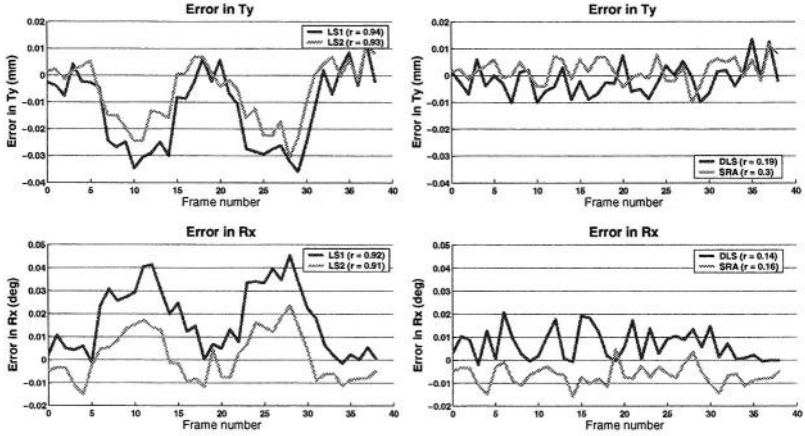


Fig. 2. Registration errors for the parameters t_y and r_x for the motion-free simulated time series. Graphs refer to LS1 and LS2 (*left*) and to DLS and SRA methods (*right*).

Results are presented in Figure 3 by subtracting the true motion from the calculated motion-correction parameters and show that both DLS and SRA methods are significantly robust to activation presence. The summary of correlation coefficients is presented in Table 1 *(right).*

param.	LS1	LS2	DLS	SRA	LS1	LS2	DLS	SRA
t_x	0.12	0.13	0.06	0.10	0.26	0.04	0.10	0.10
t_y	0.94	0.93	0.19	0.30	0.94	0.94	0.13	0.00
t_z	0.93	0.76	0.25	0.11	0.92	0.68	0.50	0.07
r_x	0.92	0.91	0.14	0.16	0.91	0.94	0.21	0.30
r_y	0.26	0.04	0.16	0.18	0.43	0.01	0.37	0.23
r_z	0.33	0.48	0.17	0.28	0.23	0.24	0.30	0.22

Table 1. Correlation values for the motionless simulated time series *(left)*, and for the simulated time series with true activation-correlated motion *(right).*

3.3 Experiments with the Real Time Series

The four registration methods were also applied to three actual time series. For these datasets, the activation profile used to compute cross-correlation was obtained by convolving the task timing with the SPM99 hemodynamic model. A moving average was removed from the estimated motion parameters before computing the correlation in order to discard slow motion trends. In the case of the DLS method, the number of activated voxels in the three mask comprised, respectively, 19%, 22% and 18% of the brain size.

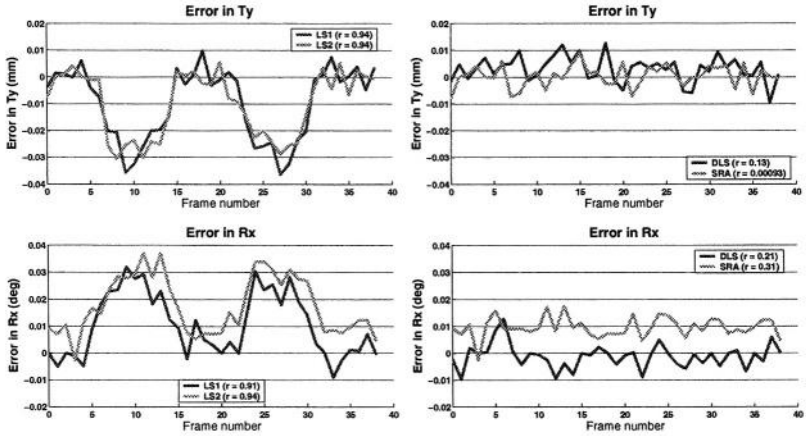


Fig. 3. Registration errors for the parameters t_y and r_x for the simulated time series with true correlated motion (true motion was removed). Graphs refer to LS1 and LS2 *(left)* and to DLS and SRA methods *(right).*

The results obtained with the three actual time series also indicate a reduction in the correlation between the motion estimates and the activation paradigm for the DLS and SRA. This is particularly visible in the t_y (and r_x) parameter (see Figure 4). Correlation values are presented in Table 2.

For the first actual time series, one can see that the different methods do not generally agree in the estimation of r_y (and r_z) parameter (see Figure 4). This effect, which is also observed for the other two time series, may be due to the fact that the methods do not share the same computational framework, as mentioned above.

param.	LS1	LS2	DLS	SRA	LS1	LS2	DLS	SRA	LS1	LS2	DLS	SRA
t_x	0.27	0.24	0.27	0.05	0.17	0.25	0.24	0.11	0.36	0.38	0.40	0.08
t_y	0.65	0.49	0.17	0.05	0.57	0.51	0.27	0.16	0.67	0.47	0.29	0.05
t_z	0.46	0.14	0.14	0.31	0.63	0.32	0.29	0.15	0.64	0.42	0.11	0.01
r_x	0.72	0.72	0.10	0.09	0.72	0.73	0.37	0.17	0.69	0.69	0.05	0.00
r_y	0.02	0.03	0.05	0.12	0.05	0.13	0.16	0.22	0.01	0.34	0.04	0.07
r_z	0.01	0.12	0.13	0.03	0.20	0.40	0.03	0.09	0.38	0.31	0.17	0.03

Table 2. Correlation values for the three real time-series.

4 Discussion

The problem of minimizing the bias introduced by the presence of activation is of particular importance due to the use of high field magnets (> 3*T*), which increase

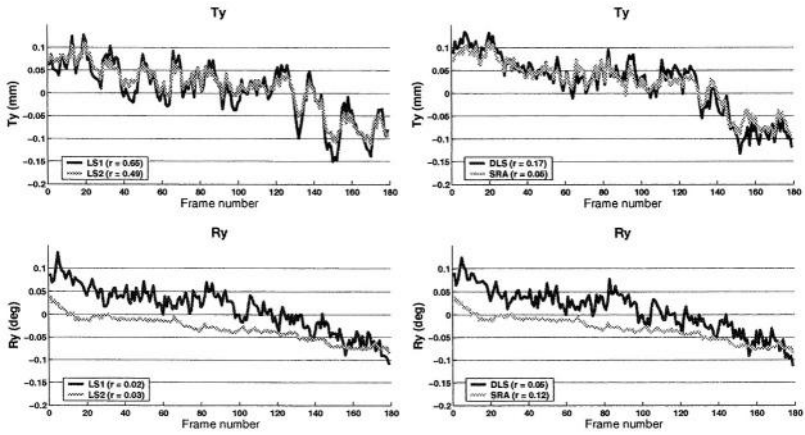


Fig. 4. Detrended registration parameters t_y and r_y for the first actual time series. Graphs refer to LS1 and LS2 *(left)* and to DLS and SRA methods *(right).*

activation amplitude. The work presented in this paper shows that the SRA and DLS methods seem suitable for the problem of motion compensation of FMRI studies, even in a situation where true activation-correlated subject motion is present. Indeed, this is an important issue when assessing the robustness of a registration method. The explanation is twofold: the first deals with the fact that interpolation errors during registration could be confounded with an activation-like bias; the second, to the well known fact that registration error (generally) increases with the initial misalignement.

The three actual time series used in this work were selected from among 14 subjects because they clearly presented a strong correlation with activation paradigm. Nevertheless, the true motion for these subjects is unknown and may or may not be correlated to the stimulus. However, the results obtained from the experiments performed in this paper clearly support the idea that the bias in motion estimates was due, at least in part, to presence of activation. Indeed, incorporating the activation profile into the SRA method or discarding about 20% of the voxels in the DLS method substantially reduces the correlation with the task.

A few plots obtained from the actual data show a disagreement between the different registration methods. In our opinion, this situation may stem from the fact that the time series include spatial distortions induced by the fast acquisition scheme. Indeed, there is an interaction between these distortions and head movement and therefore, the rigid registration approach cannot perfectly align the time series. In such ill-posed situations, similarity measures are prone to several local minima, which are due to local good matches between object features, possibly caused by the fact that both methods rely on different interpolation methods. This may explain why the two different computational frameworks

sometimes provide slightly different solutions for the rotation parameters for LS1 and LS2.

The success of the SRA method described in this paper calls for the development of integrated methods mixing motion estimation, activation detection and distortion correction. Like activation detection, however, distortion correction may require additional information, such as a magnetic field phase map obtained from the MR scan [9], adding another level of complexity because this phase map may depend on the head position in the scanner.

5 Conclusion

During the last decade, numerous general purpose similarity measures have been proposed to tackle medical image registration problems. They have led to a lot of success with important impact on neuroscience and clinical applications. The assumptions underlying these similarity measures, however, often neglect some features specific to FMRI data, leading to the kind of bias mentioned in this paper. In our opinion, tuning registration methods to account for these features, as demonstrated for the DLS and SRA methods, will be a necessary and fruitful endeavour.

References

1. L. Freire and J.-F. Mangin, "Motion correction algorithms may create spurious brain activations in the absence of subject motion," *NeuroImage,* vol. 14 Sep, pp. 709–722, 2001.
2. M. Black and A. Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications in early vision," *International Journal of Computer Vision,* vol. 19, pp. 57–91, 1996.
3. C. Nikou, F. Heitz, J.-P. Armspach, I.-J. Namer, and D. Grucker, "Registration of MR/MR and MR/SPECT brain images by fast stochastic optimization of robust voxel similarity measures," *NeuroImage,* vol. 8(1) Jul, pp. 30–43, 1998.
4. L. Freire, A. Roche, and J.-F. Mangin, "What is the best similarity measure for motion correction of fmri time series?" *IEEE Trans. on Medical Imaging,* vol. 21(5) May, pp. 470–484, 2002.
5. L. Freire and J.-F. Mangin, "Two stage alignment of fmri time series using the experiment profile to discard activation related bias," in *Medical Image Computing and Computer-Assisted Intervention,* 2002, pp. 663–670.
6. J. Orchard, C. Greif, G. Golub, B. Bjornson, and M. Atkins, "Simultaneous registration and activation detection for fMRI," *IEEE Trans. on Medical Imaging,* vol. 22(11) Nov, pp. 1427–1435, 2003.
7. M. Unser, A. Aldroubi, and M. Eden, "B-spline signal processing: Part I-theory." *IEEE Trans Signal Process,* vol. 41, pp. 821–833, 1993.
8. W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipies in C,* 2nd ed. Cambridge University Press, 1995.
9. C. Studholme, R. Constable, and J. Duncan, "Accurate alignment of functional EPI data to anatomical MRI using a physics-based distortion model," *IEEE Trans. on Medical Imaging,* vol. 19(11) Nov, pp. 1115–1127, 2000.

A Robust Algorithm for Nerve Slice Contours Correspondence

Shaorong Wang[1,2], Liyan Liu[1,2], Fucang Jia[1,2], Hua Li[1]

[1]Key Lab of Intelligent Information Processing, Institute of Computing Technology
Chinese Academy of Sciences, Beijing, 100080
{shrwang, lyliu, fcjia, lihua}@ict.ac.cn, www.ict.ac.cn
[2]Graduate School of the Chinese Academy of Sciences, Beijing 100039

Abstract. In this paper, the three-dimensional reconstruction of consecutive brachial plexus slices is discussed. To get precise results, the contour-based reconstruction method is adopted. As the contour correspondence problem is of great importance to the reconstruction, a cluster method is firstly used to get rid of the offsets between the contours introduced during the stage of data acquisition. Besides correspondence check, a robust similarity-based correspondence algorithm is proposed which results in an automatic correspondence rate more than 97%.

1 Introduction

The repair and regeneration of damaged brachial plexus has always been a focus in orthopaedics. Until now there still exists imperfectness in the restore of nerve function. The main reason lies in the wrong anastomose of nerve fibers between sense and motor ones. So it's an urgent problem in basic and clinical researches to know the exact structure of brachial plexus [16].

The structure of brachial plexus is very complicated. The nerve bundles branch, cross and recombine one another. Sense fibers and motor ones in nerve bundles mix together to combine a commix bundle in every part of brachial plexus. In this case, a three-dimensional model would provide a good understanding of the nerve structure where the traditional two-dimensional nerve atlas suffered.

In this paper, the contour-based method is introduced to reconstruct brachial plexus's three-dimensional structure, including both out contour and the ultra-complicated pathways of nerve bundles inside. A robust similarity-based contour correspondence algorithm is proposed to guarantee a precise and correct result.

2 Related Work

The volume data visualization methods can be grouped in two classes according to different problems and data characteristics [14]: surface-based reconstruction [12] and direct volume rendering [13].

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 286-293, 2004.

Surface-based reconstruction also has two categories, two-dimensional contour-based reconstruction [12] and iso-surface extracting-based reconstruction [2].

The former is a commonly used and effective visualization method. It means to reconstruct surfaces of three-dimensional objects based on a collection of planar contours representing cross-sections through the objects. It is mainly composed of three problems [3]: correspondence problem, tilling problem, and branching problem. Among them, contour correspondence, whose goal is to obtain correspondence relationship of contours on two adjacent slices, is vital to the whole progress. [5] computes the overlapping area of contours on adjacent slices as the correspondence criterion. [1],[9] approximate the contours by ellipses and then assemble them into cylinders to determine the correspondence. [4] uses domain knowledge to constrain the problem. [7] realizes automatic reconstruction through Greeb graph. [6] uses a method of distance transform and region correspondence. [8] realizes a nerve reconstruction tracing system, in which the primary rule for correspondence is to compute the overlapping area and distance between the centers of the contours on adjacent slices.

3 Nerve Reconstruction

In this paper, we realize a nerve reconstruction system to deal with 842 brachial plexus slices. In order to obtain precise three-dimensional structure of all nerve bundles and display clearly their relationship, the contour-based reconstruction method is adopted.

The whole process can be divided into several steps including image adjustment, image segment, contour correspondence, contour reconstruction and rendering. The source data is a series of two-dimensional images containing brachial plexus information. Fig.1-a is one of these slices with the thickness of 15 micron and interval of 0.2mm.

- Image adjustment: Based on the markers on the original images, the adjustment is accomplished by matching corresponding landmarks in both images through an affine transformation.
- Image segmentation: Firstly each adjusted image is binarized, and then the contours on each slice are extracted by use of watershed-based algorithm. Due to the complexity of nerve structure, automatic segmentation associated necessary human intervention is adopted.
- Contour correspondence: To find correct connection relationship between contours on adjacent slices. This is the main topic of this paper, and we will give a more detail description later.
- Contour reconstruction: [10],[11]'s methods are used here to reconstruct the three-dimensional surface structure of the brachial plexus.
- Rendering: The final stage of the process is to render the results of reconstruction on the screen.

The main contributions of this paper are listed as following:

- Cluster small contours into big circles. Perform the correspondence on the scale of big circles firstly to get rid of offset errors introduced during the data acquisition.
- Execute the correspondence progress repeatedly and in each loop put those corresponded contours as benchmarks for next correspondence.
- Check intermediate results by the constrain criterion and recorrespond the uncorresponded contours with a new method by a new set correspondence intensity.

4 Contour Correspondence

Correspondence problem is the most important and difficult problem of the contour-based reconstruction method. In our case, the structure of brachial plexus is very complicated and the offsets between the contours on the adjacent slices are sometimes very large. So the general contour-based methods suffer here.

[15] uses (1) to compute the discrepancy of two contours. If the discrepancy is lower than a given value, there exists correspondence between the two contours. Through this method an automatic correspondence rate of near 90% is obtained.

$$Disci_j = D(C_i, C_j)\frac{|C_i.Area - C_j.Area|}{\min(C_i.Area, C_j.Area)} \quad (1)$$

Careful observation of the slices shows that the contours representing nerve fibers assemble into several big circles or ellipses, which represent nerve bundles. And offsets of the fibers in the same bundles are very similar between the adjacent slices. According to above characteristics, we present our algorithm as following:

1. For each slice, cluster the small contours into several big circles.
2. Correspond the big circles, compute the corresponding circles' offsets and add them to the small contours.
3. Correspond the small contours.

4.1 Contours Cluster

Firstly we define the distance between contour C_1 and contour C_2 by (2). Suppose the total number of contours is *m*. We define a *(m+2)*-dimension vector D_k for each contour C_k by (3), in which *(x, y)* represents the center coordinate of C_k. Then cluster the vectors into several groups by hierarchical cluster method, in which the distance matrix is computed by Minkowski's metric with an exponent of 3. Necessary human interaction is needed to correct cluster errors.

$$dis(C_1, C_2) = \min\{dis(x, y) \mid x \in C_1, y \in C_2\} \quad (2)$$

$$D_k(i) = dis(C_k, C_i)(i = 0,1,..., m-1), D_k(m) = x, D_k(m+1) = y \quad (3)$$

Now the small contours have been clustered into several groups, which can be approximated by big circles, as shown in Fig.1b. The centre of each big circle is the

centroid of its corresponding group. Then for each small contour, it is approximated by a circle whose center and area is same as that of the contour. The radius of the big circle G is computed by (4).

$$G.Radius = \max\{dis(G.Center, C.Center) + C.Radius) \mid C \text{ belongs to } G\} \quad (4)$$

4.2 Big Circles Correspondence and Offset Computation

In some cases, when two contours are very far away and their areas are almost the same, the discrepancy computed by (1) may be very small, which leads to a correspondence between them. But in fact, this correspondence doesn't exist here. So, we substitute the formula (1) with a new formula (5) to compute the discrepancy. Fig.1b. shows the correspondence results of big circles. The automatic correspondence results would be nearly 100% correct as long as appropriate parameters are chosen.

$$Disc_{ij} = \sqrt{\max\{C_i.Area, C_j.Area\} / \min\{C_i.Area, C_j.Area\}}\, D(C_i, C_j) \quad (5)$$

Now compute the offsets of corresponded big circle pairs. Considerable offsets indeed exist between some circles. For each small contour, according to its subordinate group, add corresponding offset to it. As shown in Fig.1-c, the green hollow contours are processed results, compared with their counterpart on the adjacent slice (blue hollow ones). It can be seen that the error has been nicely amended. Until now, most of the error introduced during the data acquisition stage has got correction. As shown in Table. 1, the above progress does enhance the correspondence rate.

4.3 Small Contours Correspondence

This section is the most principal part of the paper. We propose a new robust algorithm to perform correspondence of complex contours, which proves to be effective and efficient. The basic idea is to complete the task from the easy to the difficult. Firstly those contours with remarkable characters are handled, then the relatively more complicated cases. The process is not finished until all small contours have been matched. In each step different criteria is adopted according to different characters of the current contours.

Firstly we assume *S1* and *S2* are two adjacent slices. Compared with a given value, contours are categorized as big contours or medium contours or small ones according to their areas. The basic algorithm is as following.

4.3.1 Basic Correspondence

Suppose that $Ov(C_2,C_1)$ represents overlap area between C_2 and C_1 and $Corr(C_1,C_2)$ represents C_1 is corresponded with C_2. *T1, T2* and *T3* are given values. Take the following steps.

```
1:For each uncorresponded big contour C2 in S2.
  C1=max contour{Ov(C2,C)|C in S1}
  if Ov(C2,C1)>T1, Corr(C2,C1)
```

```
2:For each uncorresponded medium contour C2 in S2.
  C1=min contour{Disc(C2,C)|C in S1}
  if Disc(C2,C1)<T2, Corr(C2,C1)
3:For each uncorresponded small contour C2 in S2.
  C1=min{Disc(C2,C)| uncorresponded C in S1}
  if Disc(C2,C1)<T3, Corr(C2,C1)
4:For each uncorresponded big, medium, small contour
C1 in S1, take the same operations as in step 1,2,3
```

Until now, all above operations have enabled a considerable high automatic correspondence rate.

4.3.2 Correspondence Check and One-to-One Correspondence

But there are still some correspondence errors. Take an example shown as in Fig.1-d, *A, B, M* and *N* are four medium contours with almost the same areas. Both *M* and *N* are corresponded with *B* while *A* is left alone. The reason is that $dis(M,B)<dis(M,A)$ and $dis(N,B)<dis(N,A)$. So $Disc(M,B)<Disc(M,A)$ and $Disc(N,B)<Disc(N,A)$. According to our criteria, *M, N* are both corresponded to *B*. This case would not happen to big contours.

To avoid above errors, the following is the correspondence check. For each corresponded contour *C*, $CoAr(C)$ represents the total area of all the contours corresponded with *C*. If $CoAr(C)/C.Area$ is greater than a given threshold, there exist some correspondence errors with *C*. Under this condition, remove all the correspondence relationship between *C* and its counterparts.

For each uncorresponded contour C_2 in S_2, $Co_1(C_2)$ represents the set of contours whose discrepancy is less than a given value. Suppose the size of such a set is no greater than 3, or else we only choose the very three of them with the smallest discrepancy. Combine all C_2 and its corresponding $Co_1(C_2)$ into a series of set pairs $(CS_{21}, CS_{11}),\ldots,(CS_{2k}, CS_{1k})$, so that for each two contours if they are not in the same set pair, they can't be corresponded.

Given set SS_1 and SS_2, a set correspondence *SC* from SS_2 to SS_1 is all the correspondences from contours in SS_2 to contours in SS_1. The similarity of two corresponded contours is defined by (6).

$$Sim_{ij} = \sqrt{\min(C_i.Area, C_j.Area)/\max(C_i.Area, C_j.Area)} \tag{6}$$

A position relationship constraint is used here. Suppose C_{11} and C_{12} are two adjacent contours in slice S_1, so are C_{21} and C_{22} in slice S_2. If C_{11} is corresponded with C_{21} and C_{12} is corresponded with C_{22}, the position relationship between C_{21} and C_{22} is similar to that of between C_{11} and C_{12}. To represent this principle we define *K* by (7), in which $d_{mn}=C_m.Center-C_n.Center$. If $K_{ijkl}>0$, C_k and C_l are called the supporters of (C_i, C_j). If (C_i, C_j) is the correct correspondence, there exist a lot of supporters. C_k and C_l are called the opponents If $K_{ijkl}<0$. *Opp* represents the ratio of supporters in all candidates. A correspondence intensity *CI* is defined by (8).

$$K_{ijkl} = (d_{ik} d_{jl})/\max\{|d_{ik}|^2, |d_{jl}|^2\} \tag{7}$$

$$CI_{ij} = Sim_{ij}(1 - opp) \tag{8}$$

For a given set correspondence from S_2 to S_1, the set correspondence intensity is the total correspondence intensities of all the contours in S_2. If a set correspondence *SC* is the correct set correspondence, its set correspondence intensity is the maximum of the correspondence intensities of all the set correspondence from S_2 to S_1. So the correspondence problem from S_2 to S_1 is converted into another problem, whose goal is to find a set correspondence which has the maximum set correspondence intensity.

Generally, the size of such a set is not greater than 10. And under the above condition a simple traversal algorithm is enough to solve this problem.

4.3.3 Multi-to-One Correspondence

This section mainly talks about the possibility of the correspondence between the uncorresponded medium or small contours and big contours.

For contour C_2 in S_2, suppose $Ne(C_2)$ represents the set of contours in S_1 whose nearest distance with C_2 is lower than a given value. Suppose the size of such a set is no greater than 3, or else we only choose the very three of them with the smallest distance. Then:

1. For each uncorresponded media or small contour C_2 in S_2.
   - For each contour C_1 in $Ne(C_2)$, if $max(C_1.Area, CoAr(C_1))/min(C_1.Area, CoAr(C_1))$ is greater than a given value, C_1 is erased form $Ne(C_2)$.
   - For each contour C_1 in $Ne(C_2)$, if $dis(C_1, C_2)$ is the minimum of all the distances between contour in $Ne(C_2)$ and C_2, C_1 and C_2 are considered to be corresponded.
2. For each uncorresponded media or small contour C_1 in S_1, conduct the same operations as step 1.

5 Results

The brachial plexus sample is composed of 842 slices and there are altogether 40568 contours in total. Table1 shows the comparison of the correspondence results between the methods in the literates and our method. Clearly, the method in this paper provides a more satisfied solution to the problem of the correspondence between nerve contours.

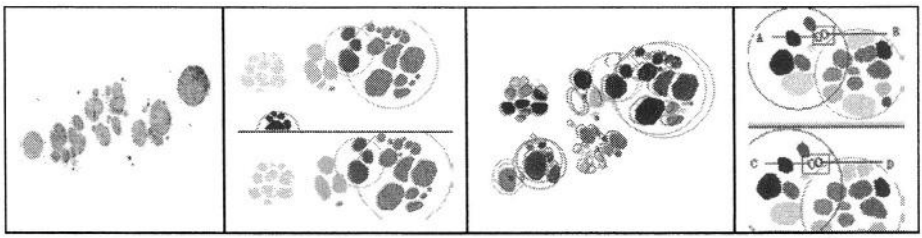


Fig. 1. a: Source image. b: Contours cluster and correspondence. c: Offsets between the adjacent slices. d: Correspondence error

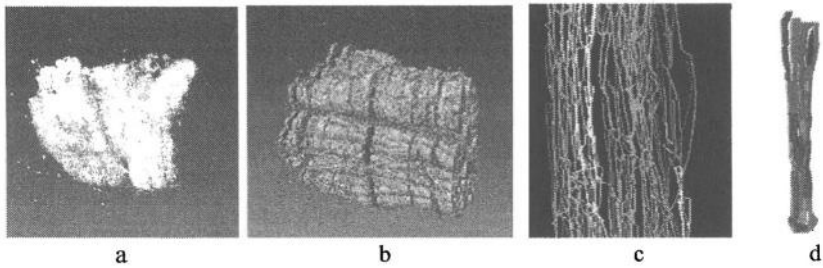


Fig. 2. Reconstruct results of different methods.

Fig.2 displays the reconstruct results of different methods. Fig.2-a is a direct volume rendering result that barely contains worthy information. Fig.2-b is obtained through Marching Cube algorithm, from which only surface information can be discerned. Fig.2-c and Fig.2-d are nerve bundles and nerve outer contours respectively, which are results of contour-based reconstruction method. They successfully realize the reappearance of the three-dimensional structure of brachial plexus including both its outer contour and the ultra-complicated pathways of nerve bundles inside, show the topographic anatomy of every nerve fascicle and its spatial relationship with nicety localization information of sense fibers and motor ones in arbitrary sections, present patterns of branching, cross and recombining of nerve bundles in the whole length. So, under the circumstances, the contour-based reconstruction behaves much better than other reconstruction methods. Now these reconstruction results have been presented to medical experts, from which they have found many significant results and applied the results into clinical brachial plexus repair [16].

Table 1. Correspondence results compared with other methods

method	contours of miscorrespondence	correct rate
Overlap	4821	88.116%
[15]'s method	4755	88.279%
Our method without offset elimination	3053	92.474%
our method	1068	97.367%

Acknowledgments

This work was funded by National Hi-Tech Research and Development (863) Plan (Grant No. 2001AA231031, 2002AA231021). National Key Basic Research Plan (grant No: G1998030608) and National Special R&D Plan for Olympic Games (grant No: 2001BA904B08). All the image data is provided by Doctor Zenggan Chen of ZhongShan Hospital of FuDan University.

References

1. D. Meyers, S. Skinner, and K. Sloan. Surfaces from contours: ACM Trans. Graphics, 11(3):228-258, Jul. 1992
2. William E. Lorenson and Harvey Cline. Two Algorithms for Three-dimensional Reconstruction of Tomograms, Medical Physics, Volume 15, Number 3, May, 1988, pp. 320-327
3. Chandrajit L. Bajaj, Edward J. Coyle, Kwun-Nan Lin. Arbitrary topology shape reconstruction from planar cross sections. Graphical Models and Image Processing 58 (1996) 524-543, 1996
4. Y. Breseler, J. A. Fessler, and A. Macovski. A Bayesian approach to reconstruction from incomplete projections of a multiple object 3D domain. IEEE Trans. Pat. Anal. Mach. Intell., 11(8):840-858, August 1989
5. Y. F. Wang and J. K. Aggarwal, "Surface reconstruction and representation of 3-D scenes." Pattern Recognition 19 (1986) 197-207
6. G. M. Treece, R. W. Prager, A. H. Gee, and L. Berman. Surface interpolation from sparse cross-sections using region correspondence. Technical Report CUED/F-INFENG/TR 342, Cambridge University Engineering Dept, March 1999
7. Y. Shinagawa and T. Kunii. Constructing a Reeb Graph Automatically from Cross Sections. IEEE Computer Graphics and Application, 11(6):45-51, November 1991
8. B.P. Burton, Automated 3D reconstruction of neuronal structures from serial sections, MS Thesis, Department of Computer Science, Texas A&M University, 1999
9. B. I. Soroka. Generalized cones from serial sections. Comp. Vis. Grap. & Image Proc., (15):154-166, 1981
10. Ganapathy S, Dennehy T G. A new General Triangulation Method for Planar Contours. Computer Graphics, 1982:16(3), pp.69-75
11. Ekoule A B, Peyrin F C, Odet C L. A Triangulation Algorithm from Arbitrary Shaped Multiple Planar Contours. ACM Transactions on Graphics, 1991:10(2), pp.182-199
12. Bernhard Geiger. Three-dimensional modeling of human organs and its application to diagnosis and surgical planning. PhD thesis, INRIA, France, 1993.
13. L. Westover, Footprint evaluation for volume rendering. Computer graphics, vol. 24, no. 4, 1990
14. T. Todd Elvins. A Survey of Algorithms for Volume Visualization. Computer Graphics, Volume 26, Number 3, August 1992, pp. 194-201
15. LiyanLiu, ShaorongWang, FucangJia, RuiChen, ShimingXiang, GuoyiLiu, HuaLi, ZhongweiChen, TongyiChen, ZengganChen, Computer-Aided Three-Dimensional Reconstruction of Microscopic Structure of Brachial Plexus Based on Consecutive Series of Slices. In proceedings of 8th international conference on CAD/Graphics pp46-51
16. ZengganChen. Computer-assisted Three-Dimensional Reconstruction of Human Brachial Plexus from Serial Tissue Sections. PhD thesis. FuDan University. China,2003

Assessing Spline-Based Multi-resolution 2D-3D Image Registration for Practical Use in Surgical Guidance

Guoyan Zheng, Xuan Zhang, Lutz-Peter Nolte*

M.E. Müller Institute for Surgical Technology and Biomechanics, University of Bern, Stauffacherstr. 78, CH-3014 Bern, Switzerland
Guoyan.Zheng@MEMcenter.unibe.ch
Tel: +41-31-631-5956; Fax: +41-31-631-5960

Abstract. A spline-based multi-resolution 2D-3D image registration algorithm has recently been introduced [1-3]. However, its accuracy, robustness, and efficiency have not been fully investigated. In this paper, we focus on assessing and improving this newly introduced 2D-3D registration algorithm. A phantom and a cadaver test, together with their respective ground truths, were specially designed for this purpose. A novel least-squares normalized pattern intensity (LSNPI) similarity measure was proposed to improve the accuracy and robustness. Several parameters that may also affect its robustness, accuracy, and efficiency are experimentally determined, including the final resolution level, the initial guess of the patient pose, the number of 2D projection images, and the angle between 2D projection images. Our experiments show that it is feasible for the assessed 2D-3D registration algorithm to achieve sub-millimeter accuracy in a realistic setup in less than two minutes, when it is used together with the newly proposed similarity measure.

1 Introduction

2D-3D registration of volumetric data with a series of calibrated and undistorted projection images has shown great potential in CT-based navigation because it obviates the invasive procedure of the conventional registration methods. In the past several years, various feature-based [4-5] as well as intensity-based [6-9] rigid 2D-3D registration algorithms have been proposed. However, registration of 2D projection images and 3D volume images is still a largely unsolved problem [9]. The main obstacles are robustness, accuracy, efficiency, and system integration [10]. A target registration error of 1-1.5 mm on average (2-3 mm worst case), plus a 95% successful registration rate on the first try, is normally required for the practical use of such an algorithm in surgical guidance [10]. Shorter running time, less ultra-operative user interaction, and fewer number of projection images required for an accurate registration are also highly appreciated in a sterilized surgical environment.

* We are thankful to S. Jonić for the spline-based multi-resolution 2D-3D image registration toolbox and insightful discussion. This research was partially funded by the Swiss National Centers of Competence in Research CO-ME.

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 294-301, 2004.

In this paper, we are interested in intensity-based 2D-3D image registration, particularly in a recently introduced spline-based multi-resolution 2D-3D registration algorithm [1-3]. It follows the same computation framework as other intensity-based 2D-3D registration algorithms. Given a set of intra-operative 2D projection images and a pre-operative 3D volume dataset, it iteratively optimizes the six rigid-body parameters describing the orientation and the translation of the patient pose, by generating digitally reconstructed radiographs (DRRs) and comparing these with the projection images using appropriate similarity measure. The difference between this algorithm and other intensity-based algorithms lies in: 1) a cubic-splines data model was used to compute the multi-resolution data pyramids, the DRRs, as well as the gradient and the Hessian of the cost function; 2) a Levenberg-Marquardt non-linear least-squares optimizer was adapted to a multi-resolution context. The registration was performed from coarsest resolution until finer one. Accuracy of approximately 1.4±0.2 mm when starting from an initial mis-registration of approximately 9.02 mm has been previously reported [3].

The work reported in this paper focuses on assessing and improving this newly introduced spline-based multi-resolution 2D-3D image registration algorithm. A novel least-squares normalized pattern intensity (LSNPI) similarity measure is proposed to improve its robustness and accuracy. Several factors that may also affect its robustness, accuracy, and efficiency are experimentally determined, including the final resolution level, the initial guess of the patient pose, the number of 2D projection images, and the angle between these 2D projection images. Assessment of these factors could provide valuable information to help clinicians improve their surgical setup and protocol for practical use of this algorithm in surgical guidance.

The remainder of this paper is organized as follows. Section 2 briefly introduces the spline-based multi-resolution 2D-3D registration algorithm. Section 3 presents our novel least-squares normalized pattern intensity similarity measure. Section 4 describes our experimental setup. Section 5 presents our experiments and the assessment results. Finally, section 6 discusses the results of our investigation.

2 Spline-Based Multi-resolution 2D-3D Image Registration

Spline-based multi-resolution 2D-3D image registration was first introduced by Jonić et al. in [1]. Here we summarize it for completeness. Details of this 2D-3D registration algorithm can be found in their previous publications [1-3].

Data Model – A continuous image model based on cubic splines was used to interpolate accurately the CT volume data as well as the 2D projection images. The advantage of using the cubic-splines data model is the possibility of having the gradient of the dissimilarity measure well defined everywhere, which is a necessary condition for accurate registration according to [3].

Synthetic Projection – A faster computation of the synthetic projection is proposed in [1] by replacing the 3D interpolation with a 2D interpolation. The principle is to adapt the sampling step for each ray casting such that only samples with an integer argument take part in the sum.

Similarity Measure – This measures the difference between the synthetic projections and associated X-ray projections. In our implementation, a novel least-squares normalized pattern intensity, which will be described in next section, is proposed to improve robustness and accuracy.

Optimization – A Marquardt-Levenberg non-linear least-squares optimization was adopted [1]. This optimization converges efficiently by considering all optimization parameters in parallel and is particularly well adapted to a multi-resolution method.

3 Least-Squares Normalized Pattern Intensity

Pattern intensity was proposed by Weese et al. [6] (see equation (1)). It measures the "smoothness" of the direct image difference $I_{diff} = I_{ref} - sI_{DRR}$ for each pixel in a small neighborhood. It considers a pixel to belong to a structure if it has a significantly different intensity value from its neighboring pixels. Penney et al. [7] have reported that pattern intensity is able to register accurately and robustly, even when soft tissues and interventional instruments are present in the fluoroscopic images. A disadvantage of pattern intensity is that there are three parameters need to be experimentally determined based on the properties of the CT and fluoroscope: constant σ which weights the function, r which defines the size of the neighborhood, and s which is a scaling factor used to build direct image difference I_{diff}.

$$S_{pi} = \sum_{q=1}^{N} \sum_{(k,l)\in D_q} \sum_{(x,y)\in(((k-x)^2+(l-y)^2))<r} \frac{\sigma^2}{\sigma^2 + (I_{diff}(x,y) - I_{diff}(k,l))^2} \quad (1)$$

where N is the number of 2D reference projection images, D_q is the region of interest in q-th projection images.

In addition, the least-squares difference (LSD) between the reference projection images and their simulations after normalizing their respective intensity ranges (as defined by equations (2)) was selected as the similarity measure in [3] for its simplicity. The advantages of this similarity measure are that its least-squares form could be well adapted for the Levenberg-Marquardt non-linear optimizer and that there is no parameter to be experimentally determined. But it might be less robust and accurate when interventional instruments are present in the fluoroscopic images.

$$I_{diff} = \frac{I_{ref} - \bar{I}_{ref}}{\sigma_{ref}} - \frac{I_{DRR} - \bar{I}_{DRR}}{\sigma_{DRR}}, \quad S_{LSD} = \frac{1}{2}\sum_{q=1}^{N} \frac{1}{card(D_q)} \sum_{(k,l)\in D_q} (I_{diff}(k,l))^2 \quad (2)$$

To improve the robustness and the accuracy and to make the algorithm more generic, a novel similarity measure called least-squares normalized pattern intensity (LSNPI) was proposed as described by equation (3), where the difference image (I_{diff}) is defined as same as in equations (2). It has the advantages present in both equation (1) and equations (2). Now only one parameter r, which defines the size of neighborhood, must be tuned.

$$S_{LSNPI} = \frac{1}{2}\sum_{q=1}^{N} \frac{1}{card(D_q)} \sum_{(k,l)\in D_q} \frac{1}{card(x,y)} \sum_{(x,y)\in(((k-x)^2+(l-y)^2))<r} (I_{diff}(x,y) - I_{diff}(k,l))^2 \quad (3)$$

4 Experimental Setup

Phantom – A rigid plastic phantom was specially designed for this study (Figure 1, left two images). It consists of one main body and four arms that are asymmetrically arranged on the body. Six titanium fiducial markers (Praxim-Medivision, Bern, Switzerland) were implanted in a way to minimize the error on computing the ground truth transformation.

Cadaver – A frozen cadaver spine specimen was prepared for this study. Five titanium fiducial markers were implanted on it (Figure 1, right two images). Note that the same data set was provided to Jonić, et al. It was previously used in [2] and [3].

Image Acquisition Protocol – Both the phantom and the cadaver were scanned by a GE LightSpeed Ultra CT scanner (GE Healthcare, Chalfont St. Giles, United Kingdom) with same intra-slice resolution (0.36 mm x 0.36 mm) but with different inter-slice thickness, 1.25 mm for the phantom and 2.5 mm for the cadaver, which resulted in volume dataset size 512x512x93 voxels for the phantom and 512x512x72 voxels for the cadaver, respectively. The 2D projection images of both the phantom and the cadaver (four images chosen from seven images for each subject, with an angular interval of 45°) were acquired from a Siemens Iso-C C-arm (Siemens AG, Erlangen, Germany). They were calibrated and undistorted with custom-made software with high accuracy [11].

Ground Truths – Both the phantom and the cadaver were equipped with infrared light emitting diode (LED) markers to establish a patient coordinate system (P-COS) and were tracked using an optoelectronic position sensor (Northern Digital OptoTrak 3020). The actual locations of fiducial markers were digitized in P-COS using an optoelectronically tracked pointer and were matched to the corresponding points in the CT volume dataset. The ground truths were then obtained using singular value decomposition with an accuracy of 0.52 mm for phantom and 0.65 mm for cadaver.

Hardware and Software – For all experiments, we used a Sun Blade 1000 (Sun Microsystems, Mountain View, CA) with 1 x UltraSparc3 600 MHz CPU and 1 GB of RAM. All programming was done using Sun CC 6.2 on SunOS 5.8; additional functionality was implemented using Qt 3.1.0 (TrollTech, Oslo, Norway).

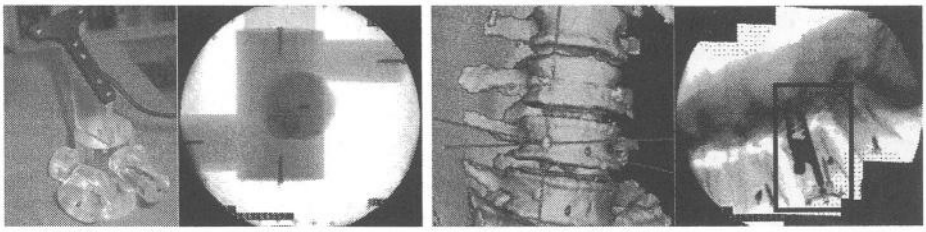


Fig. 1. Left two images: phantom (first image) and its 2D C-arm image (second image). Right two images: volume rendering result (third image, where cross lines show the position of a fiducial marker in CT space) of cadaver data with visible fiducial markers, and 2D C-arm image (forth image) with interventional instruments (delineated by black box).

5 Experiments and Assessment Results

Resolution Level – For practical use of a 2D-3D registration algorithm in surgical guidance, there should be a balance between accuracy and running time. In this experiment, we tried to investigate the relationship between the improved resolution, the running time, and the gained accuracy. This experiment was performed only on the phantom data with four C-arm images. Five levels of CT-volume pyramid as well as the same levels of C-arm image pyramid were created. Based on this investigation (see Table 1), the fourth resolution level was chosen for our later investigation.

Similarity Measure Behavior – In order to get an estimate of the behavior of the newly proposed similarity measure, cuts through the known optimum registration (ground truth) of the simulated DRRs for both the phantom (Figure 2, top two rows, rotation varies from –80° to +80° and translation from –80 mm to +80 mm around phantom ground truth) and the cadaver data (Figure 2, bottom two rows, rotation varies from –15° to +15° and translation from –15 mm to +15 mm around cadaver ground truth) were made by varying only one of the six parameters and leaving the other parameters unchanged. The same has also been done with the similarity measure used in [3], which is the least-squares difference (LSD) of normalized images as described by equations (2). It was found that there was no big difference for both similarity measures in phantom study. However, a difference in the cadaver study, when an interventional instrument is present in the C-arm images, was observed. In the figure for rotation around y-axis, a clear minimum is visible for LSNPI but not for LSD.

Initial Patient Pose – The range of the initial guess of the patient pose was determined by starting the registration from 100 uniformly distributed random positions within a range and then taking the biggest range that had a 95% successful registration rate. It was found that the range of the initial guess of the patient pose for phantom data was as much as 30° for each angular parameter and 20 mm for each translation parameter around the ground truth but only 10° and 10 mm for cadaver data.

Number of Projection Images – In this experiment, the relationship between the number of distinct projection images used for registration and the registration accuracy was investigated. The results for both the phantom and the cadaver are shown in figure 3(a). It was found that there was no direct link between the increase of the number of distinct projection images and the improved accuracy, as long as two nearly orthogonal projection images were used.

Table 1. The relationship between the resolution level, the time, and the gained accuracy

Resolution	CT Volume (vxls)	C-arm Image (pxls)	Accuracy (mm)	Running Time (s)
Initial	———	———	13.39	———
5th-level	32 x 32 x 93	48 x 36	0.95	18
4th-level	64 x 64 x 93	96 x 72	0.67	79
3rd-level	128 x 128 x 93	192 x 144	0.61	1085
2nd-level	256 x 256 x 93	384 x 288	0.58	36685
1st-level	512 x 512 x 93	768 x 576	———	Too long

Angle Between Projection Images – This experiment was performed only on phantom data, as the quality of cadaver C-arm images varies greatly from image to image. Figure 3(b) shows the results in two resolution levels. It was found that using nearly orthogonal image pair could achieve better accuracy.

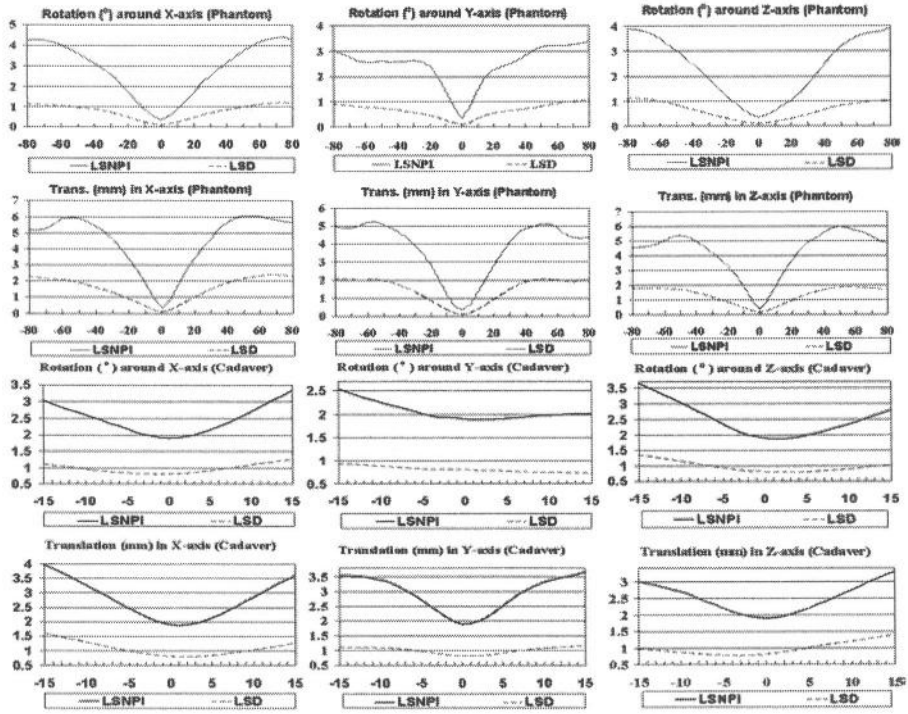


Fig. 2. Cuts through the minimum of the newly proposed least-squares normalized pattern intensity (LSNPI, solid line) and the least-squares difference of normalized images used in [3] (LSD, dash line). The ordinate shows the value of the similarity measure function, which is given as function of the orientation and translation parameters, where zero means the ground truth for each individual parameter. The top two rows show the results on the phantom data. And the results on the cadaver data are depicted by bottom two rows.

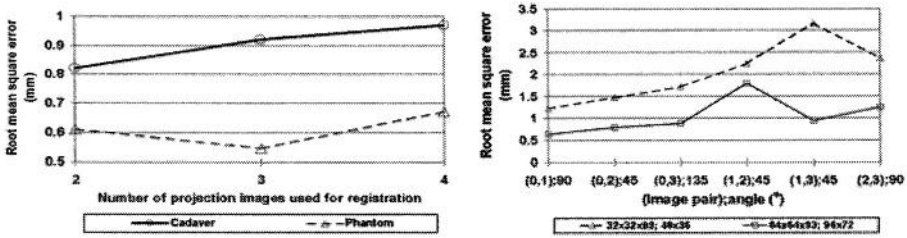


Fig. 3. (a) Left: accuracy dependence on number of the projection images; (b) Right: accuracy dependence on angle between image pair in two resolution levels.

Validation Experiment – The overall robustness, accuracy, and efficiency of the registration algorithm, together with the newly proposed similarity measure, were validated by this experiment with cadaver data. In this experiment, only two nearly orthogonal C-arm images of the cadaver were used. The algorithm stopped at the fourth resolution level. The initial patient pose was randomly created from a range of (-10°, +10°) around the ground truth for angular parameters and a range of (-10 mm, +10 mm) for translation parameters. This was repeated for 20 times. Each time the algorithm could successfully converge to the target registration error. The results are shown in Table 2. The registration error was calculated using the following equation:

$$M = \frac{1}{card(f)} \sum_{k \in f} \left\| P_k^{CT-COS} - A_s \cdot P_k^{P-COS} \right\| \tag{4}$$

Where A_s is the transformation calculated from patient pose $s = (\theta_x, \theta_y, \theta_z, t_x, t_y, t_z)$; and f is the set of fiducial markers.

Table 2. Validation experiment results. The mean and variance of initial patient pose were recorded as well as those of the registration results.

Resolution level	Deviation in Rotation (°)			Deviation in Translation (mm)			Registration Error (mm)	Running Time (s)
	$\Delta\theta_x$	$\Delta\theta_y$	$\Delta\theta_z$	Δt_x	Δt_y	Δt_z		
Initial	6.5±2.8	6.3±3.1	6.2±2.7	3.8±2.6	5.8±2.9	4.7±2.4	12.0±1.7	———
5th-level	2.2±0.8	1.6±0.6	0.9±0.5	1.4±0.5	0.6±0.2	1.5±0.5	1.5±0.1	20.5±4.6
4th-level	1.2±0.4	0.4±0.2	1.0±0.3	0.3±0.1	0.4±0.2	0.8±0.3	0.8±0.1	97.5±15.5

6 Discussion and Conclusion

We have assessed a spline-based multi-resolution 2D-3D registration algorithm through a series of experiments. Several factors which might affect its robustness, accuracy, and/or efficiency have been identified and experimentally determined. We have explored factors related to 3D volume data as well as those related to 2D projection images. To further improve its robustness and accuracy, we have proposed a novel similarity measure and validated it through experiments on phantom data and on cadaver data. The results of this investigation can help clinicians improve their surgical setup and protocol for practical use of this algorithm in a realistic environment.

The higher resolution of the data used, the more accurate the registration can be. But the higher resolution also means a longer computation time. With the continuous image model obtained from cubic splines interpolation, the algorithm could converge to an accurate solution even in low resolution.

Our newly proposed least-squares normalized pattern intensity is a similarity measure that might have the same accuracy and robustness as pattern intensity does, but it is more convenient to use and better adapted to the Levenberg-Marquardt optimizer. It was found that in a realistic setup, when interventional instruments were present in the projection images, this similarity measure was superior to a least squares difference of normalized images but with a cost of higher computation.

It was suggested by Yao et al. [12] that the more distinct X-ray images were used, the more accurate the registration could be. However, we did not find a direct link between the increase of the number of X-ray images and the improved accuracy, as long as two nearly orthogonal X-ray images were used. One possible explanation is that we only tested on up to four images. But they also found that if only two images were available, the best results were obtained when they were nearly orthogonal.

It is difficult to make strong conclusions with limited data. However, when we look at the data, we feel that we have assessed the algorithm on a data set with good quality (phantom) as well as on a data set with poor quality (cadaver). The assessed algorithm can converge to an accurate solution in both situations. The validation experiment results on the cadaver data lead us to conclude that it is feasible for this assessed 2D-3D registration algorithm to achieve sub-millimeter accuracy in less than two minutes, when it is used together with the newly proposed similarity measure.

References

1. Jonić, S., Thévenaz, P., Unser, M.: Multiresolution spline-based 3D/2D registration of CT volume and C-arm images for computer-assisted surgery. In: Sonka, M., and Hanson, K.M. (eds.): Proc. SPIE. Medical Imaging 2001: Image Processing, Vol. 4322, 1101-1109
2. Jonić, S., Thévenaz, P., Unser, M.: Multiresolution-based registration of a volume to a set of its projections. In: Sonka, M., and Fitzpatrick, J.M. (eds.): Proc. SPIE. Medical Imaging 2003: Image Processing, Vol. 5032, 1049-1052
3. Jonić, S., Thévenaz, P., Zheng, G., Nolte, L.-P., Unser, M.: Spline-based registration of a 3D CT to a set of C-arm images. IEEE T Med Imaging (submitted)
4. Hamadeh, A., Lavallée, S., Cinquin, P.: Automated 3-Dimensional Computer Tomographic and Fluoroscopic Image Registration. Comp Aid Surg 1998;3:11-19
5. Tomaževič, D., Likar, B., Slivnik, T., Pernuš, F.: 3-D/2-D Registration of CT and MR to X-Ray Images. IEEE T Med Imaging 2003;22(11):1407-1416
6. Weese, J., Buzug, T.M., Lorenz, C., Fassnacht, C.: An approach to 2D/3D registration of a vertebra in 2D x-ray fluoroscopies with 3D CT images. In: Troccaz, J., Grimson, E., Mijsges, R. (eds.): Proc. CVRMED-MRCAS 1997;119-128
7. Penney, G.P., Weese J., Little, J.A., Desmedt, P., Hill, D.L.G., Hawkes, D.J.: A comparison of similarity measures for use in 2-D-3-D medical image registration. IEEE T Med Imaging 1998;17(4):586-595
8. LaRose, D., Hayouth J., Kanade, T.: Transgraph: Interactive intensity-based 2D/3D registration of X-ray and CT data. In: Kanson, K.M., (eds.): Proc. SPIE Medical Imaging 2000: Image Processing, Vol. 3979;385-396
9. Rohlfing, T., Russakoff, D.B., Murphy, M.J., Maurer, C.R.: An intensity-based registration algorithm for probabilistic images and its application for 2-D to 3-D image registration. In: Sonka, M., Fitzpatrick, J.M., (eds.): Proc. SPIE Medical Imaging 2002: Image Processing, Vol. 4684;581-591
10. Livyatan, H., Yaniv, Z., Joskowicz, L.: Gradient-based 2-D/3-D rigid registration of fluoroscopic X-ray to CT. IEEE T Med Imaging 2003;22(11):1395-1406
11. Hofstetter, R., Slomczykowski, M., Sati, M., Nolte, L.-P.: Fluoroscopy as an imaging means for computer-assisted surgical navigation. Comp Aid Surg 1999;4:65-76
12. Yao, J., Taylor, R.: Assessing accuracy factors in deformable 2D/3D medical image registration using a statistical pelvis model. In: Proc. of the Ninth IEEE International Conference on Computer Vision (ICCV 2003):1329-1334

An Augmented Reality & Virtuality Interface for a Puncture Guidance System: Design and Validation on an Abdominal Phantom

S. Nicolau[1], J. Schmid[1], X. Pennec[2], L. Soler[1], and N. Ayache[2]

[1] IRCAD-Hopital Civil, Virtual-surg, 1 Place de l'Hopital, 67091 Strasbourg Cedex
{stephane.nicolau, jerome.schmid, luc.soler}@ircad.u-strasbg.fr
http://www.virtual-surg.com, Phone number: +33 388119095
[2] INRIA Sophia, Epidaure, 2004 Rte des Lucioles, F-06902 Sophia-Antipolis Cedex
{Stephane.Nicolau, Xavier.Pennec, Nicholas.Ayache}@sophia.inria.fr
http://www-sop.inria.fr/epidaure/Epidaure-eng.html

Abstract. In order to design an augmented reality system applied to liver punctures, we devised previously the algorithms that permit to obtain quickly an accurate patient to model registration. In this article we tackle the interface design of the system. The main constraints are the speed and accuracy with which an expert can position correctly a needle into a predefined target. Moreover, to ensure the system safety, the interface has to inform the expert when a registration failure occurs. We present here our interface that allows to fulfill the intervention requirements, by combining the two classical concepts: Augmented Reality and Augmented Virtuality. A validation, on an abdominal phantom, showed evidence that an expert can reach very accurately and quickly the predefined targets inside the phantom.

1 Introduction

Fusion of intra- or pre-operative data with the reality becomes a common tool in the fields of neurosurgery and orthopaedic surgery. This fusion enables the medical expert to see through the patient and to guide his gesture with respect to the additional information provided. Generally, the fusion is made thanks to a registration between the two reference frames in which are localized the patient and the operative data.

To design such a system, two main issues have to be tackled. Firstly, it is mandatory, for security reasons, to assess experimentally the *registration accuracy* between the two reference frames. Indeed, if the accuracy provided does not fulfill the constraints needed by the intervention, the medical expert is guided by a biased information, that can lead to dangerous gesture for the patient. Secondly, we have to evaluate the efficiency and safety of the guidance interface used by the medical expert. The interface has to allow the expert to reach the registered target with an accuracy (called here *guidance accuracy*) and a duration time compatible with the intervention constraints. Moreover, the system has to

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 302–310, 2004.

enable the expert to detect quickly any failure before and during the intervention (bad registration, incorrect tool tracking...).

Our purpose is to build an augmented reality system to guide liver punctures during interventional radiology (preliminary works are described in [9,8]). According to surgeons, the *overall accuracy* (resp. the guidance step duration) of this system has to be better than 5 mm (resp. shorter than 10 minutes) to provide significant help. In our setup, we stick radio-opaque markers on the patient skin and acquire a CT-scan of his abdomen just before the intervention. Then, an automatic 3D-reconstructions of his skin, his liver and the target is performed [10]. Two cameras (jointly calibrated) view the patient skin and a square marker attached to the needle. This marker enables to locate the needle position in the cameras reference frame. The patient is intubated during the intervention, so the volume of gas in his lungs can be controlled and monitored. Then, it is possible to fix the volume at the same value during a few seconds repetitively and to perform the needle manipulation almost in the same volume's condition than the one obtained during the preliminary CT-scan. Balter [1] and Wong [11] indicates that the mean tumor repositioning at exhalation phase in a respiratory-gated radiotherapy context is under 1 mm. Thus, it is reasonable to assume that a rigid registration of the markers, visible in both CT and video images, is sufficient to register accurately the 3D-model extracted from the CT in the cameras reference frame. A quantitative validation study on a phantom, carried out in [9], showed that a mean registration accuracy σ_r of 2 mm (RMS) was reached within the liver.

In this paper, we focus on the interface design of our system, devoted to percutaneous liver punctures. In our context, knowing that $\sigma_r = 2$ mm, we can afford at most a guidance accuracy of $\sqrt{5^2 - \sigma_r^2} \simeq 4.5$ mm in order to reach the 5 mm of overall accuracy. In addition, we need a quick targeting guidance (shorter than the 10 minutes routinely needed for this kind of intervention). Eventually, the software has to enable the expert to check quickly the correctness of the model registration. Classically, there are two types of interface used in existing medical computer-aided systems. One type, so called Augmented Reality, superimposes intra- or pre-operative data on an image of the reality [4,3]. The other type, called Augmented Virtuality, displays the tool position in the reference frame of the operative data [6]. We argue in Sec. 3 that each of them presents individually advantages and drawbacks, and that an interface integrating both approaches will provide the best efficiency.

In the sequel, we first recall in Sec. 2 how we register automatically the reconstructed model and how we find in real time the needle location in the camera frame. Then, we present our interface, and we show in Sec. 3 how the double approach allows us to obtain an excellent accuracy and to secure the system during the intervention.

2 Principles of Our Guidance System

The overall purpose of our system is to guide the needle manipulated by the expert toward a predefined target. This section deals with the first steps: the computation of the transformation T relating the operative data to the camera frame, and the localization of the needle. To find T, we use the fiducials that are automatically extracted from the CT and the video images. After a matching process, a 3D/2D point-based registration is performed to relate the model and the patient in the same reference frame.

2.1 Automated Localization and Matching of Markers

The principle of the marker localization in the video images is based on a HSV color analysis, followed by a component size and shape thresholding, and the assumption that the skin takes up the main surface. The markers in the CT-image are extracted by a top-hat characterization that emphasizes small singularities on the skin surface.

The matching between the video markers is realized thanks to epipolar geometry, and, the correspondences between video and CT markers is carried out by a prediction/verification algorithm. A validation carried out in [8] showed that these algorithms are robust and that the overall computation time of the extraction, matching and registration process is below 120 sec.

2.2 Registration of the Virtual Model in the Cameras Frame

We choose a 3D/2D points registration approach to provide the rigid transformation that relates scanner frame and cameras frame. The classical choice is to optimize the SPPC criterion (see [9]):

$$SPPC(T) = \sum_{k=1}^{S} \sum_{i=1}^{N} \xi_i^k \cdot \frac{\| \tilde{m}_i^{(k)} - P^{(k)}(T \star \tilde{M}_i) \|^2}{2 \cdot {\sigma_{2D}}^2}$$

where S (resp. N) is the number of cameras (resp. markers), $\tilde{m}_i^{(k)}$ is the observed 2D coordinates of the i^{th} markers in the k^{th} video image, M_i is the observed 3D coordinates of the i^{th} markers in the CT-image, $P^{(k)}$ the projective function, ξ_i^k is a binary variable equal to 1 if the i^{th} marker is visible in the k^{th} video image and 0 if not, and T the seeked transformation. However, this criterion considers that noise only corrupts the 2D data and that 3D data are exact. In our context, this assumption is erroneous as the markers extraction from the CT-image is corrupted by noise as well.

A more realistic statistical hypothesis is that we are measuring noisy versions $\tilde{M}_i$ of the unknown exact 3D points M_i (more details are given in [9]). Moreover, we can now safely assume that all 2D and 3D measurements are independent. A

ML estimation of the transformation T and the *auxiliary variables* M_i leads to minimize the *Extended Projective Points Criterion* (EPPC):

$$EPPC(T, M_1, \ldots M_N) = \sum_{i=1}^{N} \frac{\| \tilde{M}_i - M_i \|^2}{2 \cdot {\sigma_{3D}}^2} + \sum_{k=1}^{S} \sum_{i=1}^{N} \xi_i^k \cdot \frac{\| \tilde{m}_i^{(k)} - m_i^{(k)} \|^2}{2 \cdot {\sigma_{2D}}^2}$$

The minimization procedure is consequently modified into an alternated minimization w.r.t. the seeked transformation T, and w.r.t. the M_i.

2.3 Needle Tracking

We have to track the needle location and orientation in the camera reference frame. To realize it, we attach an oriented square marker whose corners are automatically localized on video images in real-time using an adapted version of the ARTkit library [5]. Then, knowing the size of the square, we are able to localize it in the camera reference frame by minimizing the classical 3D/2D SPPC criterion. Calibrating the relative needle position w.r.t. the square marker with the pivot method [7], we are finally able to superimpose the virtual model on the real one on video images.

3 A Secured and Ergonomic Guidance Interface

Our interface has to be designed and adapted for our particular application: liver punctures. In the field of craniotomy, Grimson *et al* [4] superimpose the reconstructed model on external video image of the patient skull. This approach allows the surgeon to check instantly the validity of the registration: if the registration is false, the superimposition will be visually incorrect. However, this kind of interface provides a view that does not correspond to the surgeon natural field of view. Realized and visualized movements can be inverted. Therefore, it needs an important interpretation effort. Moreover, since the focal lengths of the cameras are fixed, no zoom of the area of interest is available.

In the context of laparoscopy guidance, Lango [6] registers the 3D reconstructed model with the patient by pointing with a tracked tool (Polaris™) several radio-opaque markers stick on the patient skin. He proposed an interface that showed the tool position with respect to the model. Moreover, he displays the 3 CT-slices where the tip of the laparoscope lies. This approach is very useful to understand the relative position of the tool with respect to the model, since the user can choose his angle of view and an appropriated zoom. Nevertheless, since there is no camera, it is not possible to display the 3D model on an external video view of the patient. Then, the quality of the registration cannot be assessed quickly during the intervention. Indeed, this can only be done interactively, at a given time point, by pointing some reference points on the patient skin. Therefore, if the patient moves after the registration has been done, it will undergo a bias. This analysis lead us to realize an interface that provides the information of both approaches.

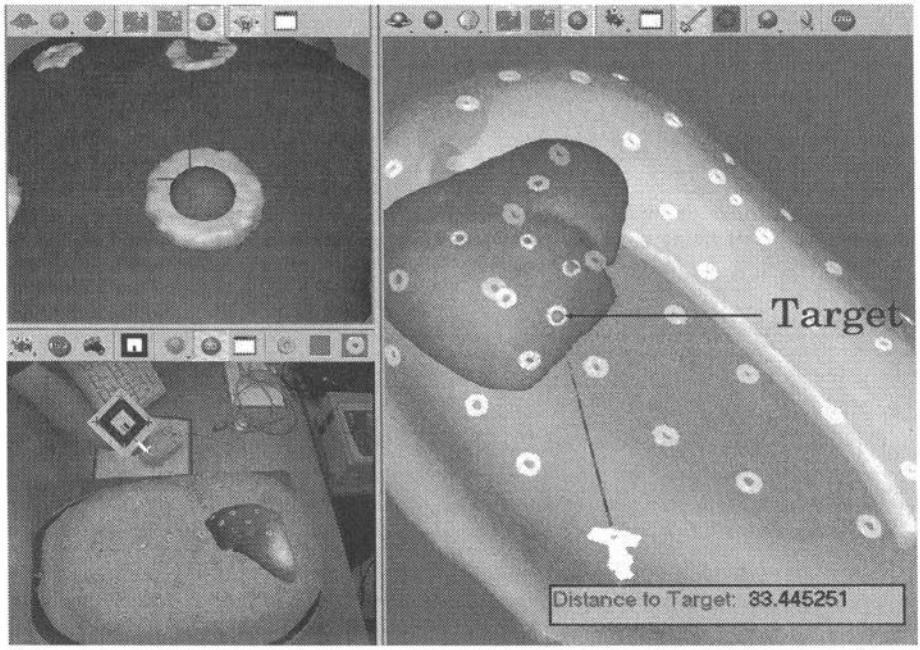


Fig. 1. Three screens guidance interface. The bottom left image corresponds to the augmented reality view, in which are displayed the 3D reconstruction of the liver and the virtual needle. The top left image displays the virtual needle view (oriented toward a marker stick on the liver surface). The right image shows the main virtual view, in which one can see the relative position of the needle w.r.t. the phantom. We indicate in its corner the virtual distance in mm that separates the tip needle to the target (in this case, a marker center).

3.1 A Three Screens Interface

Our interface (showed on Fig. 1) is divided into three screens described below. Their features and properties have been optimized with surgeons, in order to provide them a clear and intuitive tool. Each of the action associated to each screen can be done by another operator with a mouse action only (no keyboard action). These considerations should reduce time consuming manipulation.

The Augmented Reality View (Bottom Left Image in Fig. 1) In this screen, one of the two video images returned by our cameras is displayed. The user can switch between both views, enable or disable the real time superimposition of the 3D model on the video images, choose the transparency level of its different elements and display the real-time extraction of the markers. Furthermore, the user can superimpose the virtual needle on the tracked real needle and monitor the real-time tracking of the square marker attached on it. Finally,

the user can check visually the registration quality by superimposing virtual elements. If he considers that this is not acceptable (which can occur if the patient has slightly moved during the intervention), a new extraction of the markers is done in order to update the registration.

The Virtual Needle View (Top Left Image in Fig. 1) In order to direct a tool toward a target, Carrat *et al* [2] proposed three crosses displayed on a screen, that have to be superimposed. The optimal trajectory is represented by a static central cross-hair. The tool tip and axis are projected dynamically on a view orthogonal to this trajectory, and are represented by two different cross-hairs. Although this interface enables the user to reach a correct orientation, it is not very intuitive as the user looses any representation of the reality.

In the virtual needle screen, we propose to display a view that corresponds to what would see a camera positioned on the tip needle and oriented along its axis. This view was created to facilitate the orientation of the needle toward the target point. In our interface, it is represented by a green sphere of 2 mm of diameter. This view is easily understood by surgeons since it is very similar to an endoscopic view they are used to. To keep a good visibility when the needle goes through organs, the classical actions of 3D model visibility and transparency are available.

The Virtual Exterior View (Right Image in Fig. 1) In this screen, the 3D virtual scene, composed by the 3D reconstruction and the tool representation, is rendered from a viewpoint controlled by the user. Like in a classical viewer, he can rotate, translate and zoom the elements and define their properties (visibility and transparency). Moreover, it is possible to display as well the CT-scan from which the reconstruction is made, and navigate through its slices. The contrast can be enhanced like in the usual radiological viewer. When the 3D reconstructions of the liver and tumors are available the medical expert guides the needle to the tumor center, using the 3D visualization of the tumor and tip needle relative position. If the reconstructions are not available, for time or technical reasons, the expert can visualize the 3D CT-slices instead of the 3D reconstruction. Then, he can define the target position on a specific CT slice by a mouse click (cf. Fig. 2). Since it is difficult to assess visually the distance between the tip needle and the target, we print it inside the virtual exterior view.

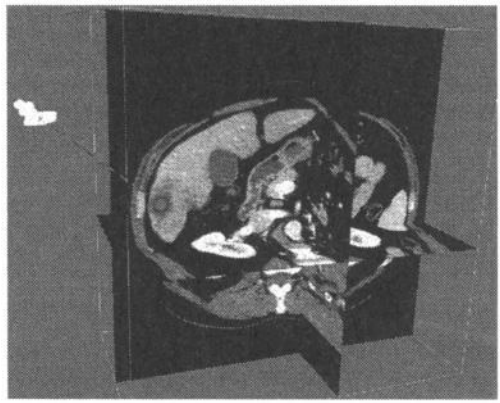


Fig. 2. Patient CT image displayed in the virtual exterior view. One can see a green sphere target that was put by the user.

3.2 Evaluation of the Overall System

The purpose of the experiment was to assess the accuracy of the needle targeting obtained by several surgeons and engineers, using our AR guidance system. Four targets were modeled with radio-opaque markers stuck on the fake liver inside the phantom. Seven participants each performed 10 consecutive needle targetings of the model tumors (cf. Fig. 3 a). During the positioning, the operator placed the needle and stopped his movement when he thought that he had reached the tumor center. After each trial, the time required to position the needle was recorded, and the accuracy of the hits was verified by an independent observer using an endoscopic camera introduced into the phantom and focusing on the targets (cf. Fig. 3 b). Accuracy and time results are shown in Table 1.

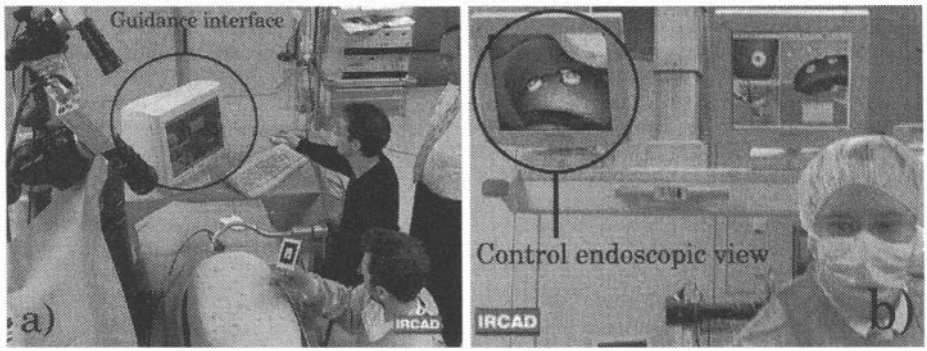


Fig. 3. a) Setup of the experiment: the user is positioning the needle, tracked by a stereoscopic system, thanks to the guidance interface. b) An endoscopic view is displayed behind the user. It enabled to assess visually the correctness of each needle targeting.

3.3 An Intuitive and Powerful Interface

The results indicate that the worst average accuracy obtained is below 3 mm, which clearly fulfills our accuracy constraint (5 mm). In addition, the system allows to reach the target very quickly (average time under 30 sec.) with respect to the usual time needed for a standard percutaneous intervention (10 minutes).

A previous experiment (see [8]), in which the user was guided only by an augmented reality screen, provided less accurate results, and more importantly longer manipulation times. It confirms the fact that the information complementarity given by the three different screens is a powerful aspect of our interface.

It has to be noticed that each one of the three screens was used intuitively at the same stage during the needle positioning by each person involved in our experiment. The *augmented reality view* has been used at the beginning of the

	Average distance (mm) ± std.	Minimum distance	Maximum distance	Average time (sec.) ± std.
Engineer 1	0.95 ± 0.67	0	2	25 ± 5.6
Engineer 2	1.7 ± 0.97	0	3	14 ± 2.0
Engineer 3	1.8 ± 0.84	0	3	18 ± 5.5
Surgeon 1	2.2 ± 0.57	1	3	32 ± 12.2
Surgeon 2	2.9 ± 1.25	0	5	22 ± 3.1
Surgeon 3	1.3 ± 1.16	0	3	32 ± 3.7
Professor	0.84 ± 0.48	0	1	32 ± 6.4
All	1.8 ± 0.7	-	-	23.8 ±7.3

Table 1. Accuracy and time results obtained by each user. The average distance, which is always below 3 mm, fulfills our accuracy constraints (5 mm). Moreover, the time needed is, by far, shorter than 1 minute, whereas an expert needs routinely 10 minutes for such intervention.

needle insertion. Firstly, it was used to check the automatic skin fiducials detection, the visual quality of the skin registration, and the tool superimposition. Secondly, it allowed to define a rough estimation of a correct skin entry point and needle orientation. During the insertion, the *virtual needle view* was always used. Indeed, it seems really adapted to needle orientation problem, since the user has only to keep the target under the cross displayed on the view: this act seemed very intuitive to everybody. Finally, the user swapped his attention to the *virtual exterior view* when the tip needle was very close to the target (below 3 mm). At this moment, a little variation of the needle position produces a big virtual view displacement. As it could make disappear the target from the *virtual needle view,* each user carried out the fine positioning with the *virtual exterior view.* At this step, he was helped by another operator that zoomed on the interest zone.

4 Conclusion

In order to design an augmented reality system devoted to liver punctures, we developed in [9,8] the procedures that allow to register accurately and quickly a patient CT model to video images. The present article deals with the interface design of our system. To overcome the constrains of this intervention (overall targeting accuracy below 5 mm, and guidance duration shorter than 10 minutes), the interface has to enable the expert to reach quickly and accurately the predefined target. Moreover, to ensure the system safety, it has to provide the expert the possibility to check visually the model registration quality during the intervention.

To fulfill these requirements, we propose a three screens interface. Its main advantage over classical augmented reality system, is that it provides two complementary kind of view: a view of the reality on which are superimposed the

patient 3D model and the virtual needle, and a virtual view of the 3D model, in which is displayed the current needle position.

This double approach enables the expert to check continuously the model registration quality, and to choose the best angle of view during the needle insertion. A validation experiment on an abdomen phantom, realized with both engineers and surgeons, proved that our interface is very intuitive and permits the user to reach the planned targets with an excellent accuracy with respect to the intervention requirements. Moreover, the average time needed for a correct needle positioning is by far smaller than the routinely intervention duration (less than 40 sec. against 10 minutes).

In the immediate future, we plan to carry out our first validation on a patient. In addition, we will adapt the current system to laparoscopic interventions. Our interface will optimize the laparoscopic tool positioning before the intervention, and it will help the surgeon by merging the 3D patient model into the endoscopic video image.

References

1. J.M. Baiter, K.L. Lam, C.J. McGinn, T.S. Lawrence, and R.K. Ten Haken. Improvement of CT-based treatment-planning models of abdominals targets using static exhale imaging. *Int. J. Radiation Oncology Biol. Phys.,* 41(4):939–943, 1998.
2. L. Carrat, J. Tonetti, P. Merloz, and J. Troccaz. Percutaneous computer-assisted iliosacral screwing: Clinical validation. In Springer Verlag, editor, *MICCAI'00,* volume LNCS 1935, pages 1229–1237, 2000.
3. J. Feldmar, N. Ayache, and F. Betting. 3d-2d projective registration of free-form curves *and* surfaces. *Journal of Comp. Vis. and Im. Under.,* 65(3):403–424, 1997.
4. W. Grimson, G. Ettinger, S. White, W. Wells T. Lozano-Perez, and R. Kikinis. An automatic registration method for frameless stereotaxy, image-guided surgery and enhanced reality visualization. *IEEE TMI,* 15(2):129–140, April 1996.
5. Hiro. Human interface technology laboratory, http://www.hitl.washington.edu/.
6. T. Lango, B. Ystgaard, G. Tangen, T. Hernes, and R. Marvik. Feasibility of 3d navigation in laparoscopic surgery., Oral presentation at the SMIT (Society for Medical Innovation and Technology) Conference. September 2002. Oslo. Norway.
7. S. Lavalle, P. Cinquin, and J. Troccaz. *Computer Integrated Surgery and Therapy: State of the Art,* chapter 10, pages 239–310. IS Press, Amsterdam, NL, in C. Roux and J.L. Coatrieux edition, 1997.
8. S. Nicolau et al. An Augmented reality system to guide radio-frequency tumor ablation. In *Journal of Computer Animation and Virtual World,* 2004. In Press.
9. S. Nicolau, X. Pennec, L. Soler, and N. Ayache. An accuracy certified augmented reality system for therapy guidance. In *European Conference on Computer Vision (ECCV'04),* LNCS 3023, pages 79–91. Springer-Verlag, 2004.
10. L. Soler et al. Fully automatic anatomical, pathological, and functional segmentation from CT scans for hepatic surgery. *Comp. Aided Surg.,* 6(3), Aug. 2001.
11. J. Wong et al. The use of active breathing control (abc) to reduce margin for breathing motion. *Int. J. Radiation Oncology Biol. Phys.,* 44(4):911–919, 1999.

Gaze Contingent Depth Recovery and Motion Stabilisation for Minimally Invasive Robotic Surgery

George P. Mylonas, Ara Darzi, Guang-Zhong Yang

Royal Society/Wolfson Medical Image Computing Laboratory
Imperial College London,
London, United Kingdom
{george.mylonas, a.darzi, g.z.yang}@imperial.ac.uk

Abstract. The introduction of surgical robots in minimally invasive surgery has allowed enhanced manual dexterity through the use of microprocessor controlled mechanical wrists. They permit the use of motion scaling for reducing gross hand movements and the performance of micro-scale tasks that are otherwise not possible. The high degree of freedom offered by robotic surgery, however, can introduce the problems of complex instrument control and hand-eye coordination. The purpose of this project is to investigate the use of real-time binocular eye tracking for empowering the robots with human vision using knowledge acquired in situ, thus simplifying, as well as enhancing, robotic control in surgery. By utilizing the close relationship between the horizontal disparity and the depth perception, varying with the viewing distance, we demonstrate how vergence can be effectively used for recovering 3D depth at the fixation points and further be used for adaptive motion stabilization during surgery. A dedicated stereo viewer and eye tracking system has been developed and experimental results involving normal subjects viewing real as well as synthetic scene are presented. Detailed quantitative analysis demonstrates the strength and potential value of the method.
Keywords: binocular eyetracking, minimally invasive robotic surgery, gaze contingent control, eye-hand coordination

1 Introduction

The field of surgery is entering a phase of continuous improvement, driven by recent advances in surgical technology and the quest for minimising invasiveness and patient trauma during surgical procedures. Medical robotics and computer-assisted surgery are new and promising fields of study, which aim to augment the capabilities of surgeons by taking the best from robots and humans. With robotically assisted minimally invasive surgery, dexterity is enhanced by microprocessor controlled mechanical wrists, which allow motion scaling for reducing gross hand movements and the performance of micro-scale tasks otherwise not possible. Current robotic systems allow the surgeon to operate while seated at the console viewing a magnified stereo image of the surgical field. His hand-wrist manoeuvres are then seamlessly translated into precise, real-time movements of the surgical instruments inside the patient. The continuing evolution of the technology including force feedback and virtual immobilization through real-time motion adaptation will permit more complex

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 311-319, 2004.

procedures such as beating heart surgeries to be carried out under a static frame-of-reference. The use of robotically assisted minimally invasive surgery provides an ideal environment for integrating pre-operative data of the patient for performing image guided surgery and active constraint control, all conducted without the need of the surgeon to remove his/her eyes from the operating field of view.

The high degree of freedom offered by robotic surgery can introduce the problems of complex instrument control and hand-eye coordination. The purpose of this paper is to investigate the use of eye gaze for simplifying, as well as enhancing, robotic control in surgery. Compared to the use of other input channels, eye gaze is the only input modality that implicitly carries information on the focus of the user's attention at a specific point in time. This research extends our existing experience in real-time eye tracking and saccadic eye movement analysis for investigating gaze contingent issues that are specific to robotic control in surgery. One key advantage of using gaze contingent control is that it allows seamless integration of motion compensation for complex motion of the soft tissue, as in this case we only need to accurately track velocity fields within a relatively small area that is directly under foveal vision. Simple rigid body motion of the camera can therefore be used to provide a perceptually stable operating field-of-view.

2 Method

2.1 Vergence as a Means for Gaze Contingent Control

One of the strongest depth cues available to human is the horizontal disparity that exists between the two retinal images. There is a close relationship between the horizontal disparity and the depth perception, varying with the viewing distance. More specifically, as the fixation point moves away from the observer, the horizontal disparity between the two retinal images is diminished and vice-versa, as illustrated in Fig. 1a. In order to extract quantitative information regarding the depth of the fixation point, ocular vergence which provides a veridical interpretation of stereoscopic depth needs to be measured [1]. One method of achieving this is to measure the corneal reflection from a fixed light source in relation to the position of the pupil. When infrared light is shone onto the eye, several reflections occur on the boundaries of the lens and cornea, the so-called *Purkinje images.* The first Purkinje image is the light reflection from the corneal bulge, often referred to as the "glint". Also, when infrared light shines the eye, the relatively dark iris becomes bright. The pupil that sinks the infrared light remains dark producing high contrast with the iris. With image processing, the centre of both the dark pupil and the glint can be identified and localized [2]. The two centres define a vector, which can be mapped to a unique eye gaze direction. This is possible because of the fact that the absolute translation of the glint and the pupil are different, as the centres of curvature of the corneal bulge and the eye are different. Since the radius of curvature of the cornea is smaller than that of the eye, during saccade the corneal reflection moves in the direction of eye movement but only about half as far as the pupil moves. The combined tracking of both eyes gaze direction, can provide the ocular vergence measure.

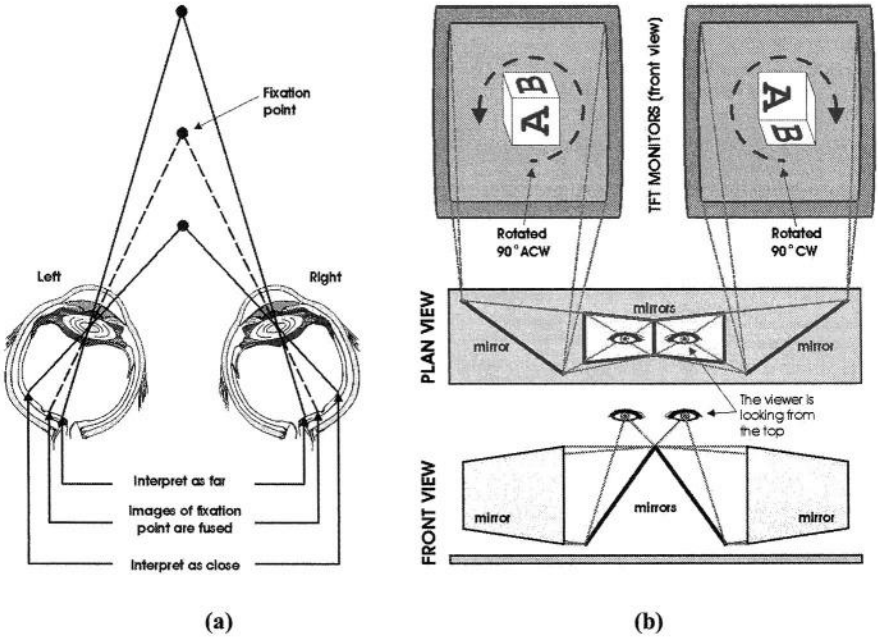


Fig. 1. (a) The relationship of the horizontal disparity between the two retinal images and depth perception, varies with the viewing distance. (b) Simplified schematic of the stereoscopic viewer with binocular eye tracking capability. Whilst the eyes being tracked, an individual fuses the two stereo images displayed on the monitors. His/her ocular vergence is determined hence the fixation point in 3D can be determined

2.2 Experiment Design and System Setup

Based on the principle described above, a stereoscopic viewer with binocular eye tracking was constructed. The optical path allowing stereo viewing is illustrated in Fig. 1b where two TFT monitors are used to display live video feed. The purpose of the mirrors used is to scale down the images in a size that matches the inter-pupilary distance, thus facilitating the fusion of the two views into a single 3D image. By using two eye tracking cameras built into the stereoscope we can quantify the ocular vergence and hence determine the depth of the fixation point. Since the two gaze vectors are expected to be epipolar, we can also determine the fixation point on the horizontal and vertical axis. In order to establish the relationship between pupil-glint vectors and points in 3D space, and to correct for subject specific variations of the eye geometry, calibration is required prior to any eye tracking session.

2.3 3D Eye Tracking Calibration

For this study, the calibration for the corneal reflection and pupil centre vectors is calculated using radial basis spline [3]. Let $\mathbf{X} = (X^1, X^2, X^3)$ be a 1-to-1 3D vector value function of a 4D eye coordinate vector $\mathbf{p} = (p^1, p^2, p^3, p^4)$. Assuming further that

vector function **X** can be decoupled into three independent scalar functions of vector **p**, each scalar component of function **X** can thus be continuously interpolated with radial basis spline, that is:

$$X^d(\mathbf{p}) = \sum_{i=1}^{N} b_i \Phi_i(\mathbf{p}) + \mathbf{a} \cdot [\mathbf{p} \quad 1]^T \tag{1}$$

where b_i is the radial basis coefficient of the corresponding basis function Φ_i and vector **a** is a global affine coefficient. The spline parameters **a** and **b** were determined by solving the following system of linear equations:

$$\begin{bmatrix} \Phi_1(\mathbf{p}_1) & \cdots & \Phi_N(\mathbf{p}_1) & & \\ \vdots & & \vdots & \mathbf{P} & \\ \Phi_1(\mathbf{p}_N) & \cdots & \Phi_N(\mathbf{p}_N) & & \\ & \mathbf{P}^T & & \mathbf{0} & \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_N \\ a^1 \\ \vdots \\ a^5 \end{bmatrix} = \begin{bmatrix} X_1^d \\ \vdots \\ X_N^d \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{P} = \begin{bmatrix} 1 & p_1^1 & p_1^2 & p_1^3 & p_1^4 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & p_N^1 & p_N^2 & p_N^3 & p_N^4 \end{bmatrix}^T \tag{2}$$

The radial basis function Φ_i was defined as that of the Euclidean distance in 4D from a given point **p** to the i^{th} control point $\mathbf{p}^i$, *i.e.,*

$$\Phi_i(\mathbf{p}) = r_i^2 \ln r_i^2 \quad \textit{where} \quad r_i = \|\mathbf{p} - \mathbf{p}_i\| \tag{3}$$

The necessary condition for the coefficient matrix in (2) not being ill conditioned is that any two given points in the 4D space are not co-planar. This criterion can be ensured by sampling the *a priori* spline control point value, (X^1, X^2, X^3), as a unique point in the calibrated 3D volume. We perform this by presenting the observer with 27 dots in sets of 9, each set displayed in each of the three calibration planes each with different depth.

2.4 Gaze Contingent Depth Recovery and Motion Stabilisation

In order to demonstrate the practical value of the proposed concept, two experiments were conducted; one involves the use of vergence for gaze contingent depth recovery for soft tissue. The other is used for adaptively changing the position of the camera to cancel out cyclic motion of a tissue in such a way that the foveal field of view is stabilized. For depth recovery, both real scenes captured by live stereo camera and computer generated surfaces are used. Five subjects where asked to observe the two images by following a suggested fixation path. Their fixation points where acquired from the eye tracker while performing the experiment and the acquired depth coordinates where recorded.

For motion stabilization, we demonstrate how gaze contingency can be used to stabilize the apparent position of a moving target by accordingly controlling the compensatory movement of the stereo camera. An OpenGL stereo scene is set up in perspective projection and a target sphere is eccentrically oscillated in the Z-axis

(depth) by keeping its X-axis and Y-axis stable. The virtual target is oscillated by transformation of the model matrix. Any required movement of the camera is respectively simulated by appropriate transformation of the view matrix. Free sinusoidal oscillation was used in this experiment. To regulate the movement of the camera, a closed feedback loop as shown in Fig. 2 was used. As the user fixates on the target, its fixation amplitude is determined and subtracted from the preset amplitude, which corresponds to the reference distance between the camera and the target that we attempt to maintain constant. The positional shift of the virtual camera is dependent on the error signal fed into the camera controller. An error of zero leaves the position of the camera unaffected while a negative or positive error shifts the camera closer or further away from the target. The response of the camera controller is regulated by Proportional, Integral and Derivative (PID) gains [4]. For this experiment the required reference distance was set to a value of +1. This means that the camera has to be kept at a constant distance of one depth units on top of the target, which in this experiment is set to sinusoidal oscillation with a frequency of 0.1 Hz.

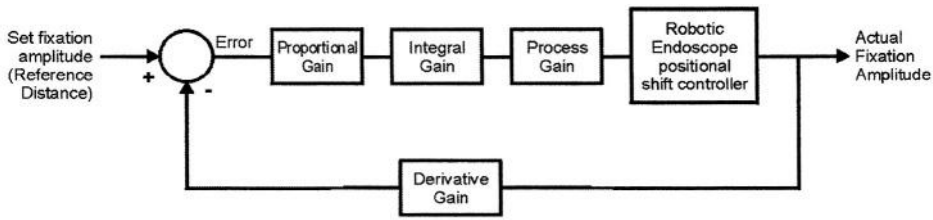


Fig. 2. By implementing a closed feedback loop, the robotic controller will shift the virtual camera in an attempt to maintain the fixation amplitude error signal down to zero

3 Results

Fig. 3 illustrates the depths recovered by the five subjects studied. During the experiment, they were asked to scan with their eyes along the Y-axis of the object. During the experiment, no visual markers were introduced and they were relatively free to select and fixate on image features of their preference Fig. 3a shows a comparative plot of the surface depths recovered from these subjects. It is evident that a relatively close correlation has been achieved, demonstrating the feasibility of veridical reconstruction of the real depth.

The same subjects were also asked to perform a similar task with synthetic images. This is necessary since in the master control console of a robotic system both live video and synthetically generated images are often present. Thus, it is important to establish that similar depth reconstruction behaviour can be achieved. Similarly to the previous experiment, the subjects were asked to follow a predefined path by fixating at image features of their preference. The only restriction posed was that they had to fixate at certain "landmark" areas, which correspond to either valleys or hills of

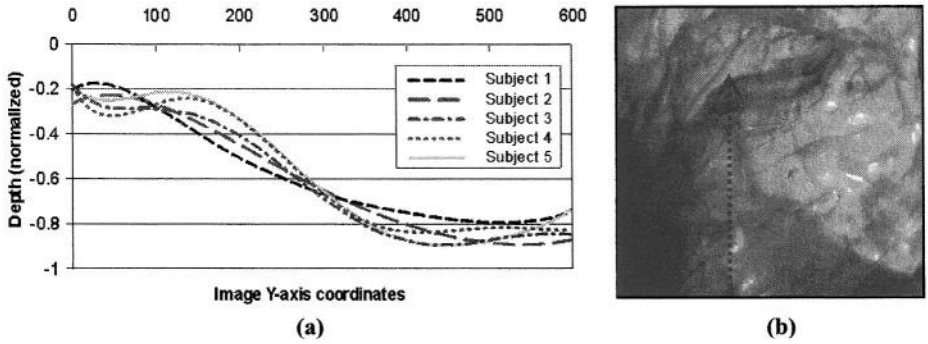


Fig. 3. (a) The comparative results of the reconstructed depths from the fixation paths of five subjects along with the surface used (b). The subjects followed a loosely suggested vertical path starting from the bottom of the surface moving towards the top.

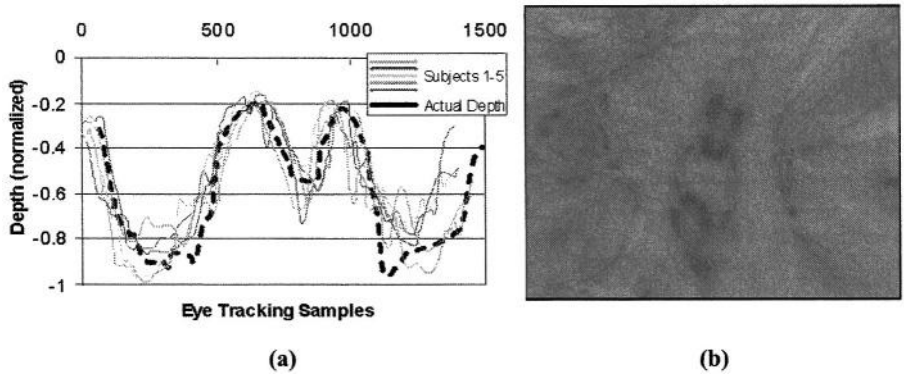


Fig. 4. (a) Graphical comparison of each subject's recovered depths against the actual depth of the suggested path, along with the virtual surface depicted on the right (b).

known depth. Fig. 4 presents all the depths recovered from these subjects. The known depth is presented as a thick line as reference.

For motion stabilization, the subjects were instructed to keep fixating on the moving target, which will become stationary after stabilization. After a short period of adaptation, all the subjects were able to stabilize the target, and Fig. 5 demonstrates the constant distance between the target and the camera that the observers were able to maintain. It is evident that the gaze contingent camera closely compensates for the oscillation of the target. To allow for a more quantitative analysis, Table 1 illustrates the regression ratios of the target and motion compensated camera position after subtracting out the constant distance maintained. The mean and standard deviation of the regression ratio achieved for this study group is 0.103 and 0.0912 respectively.

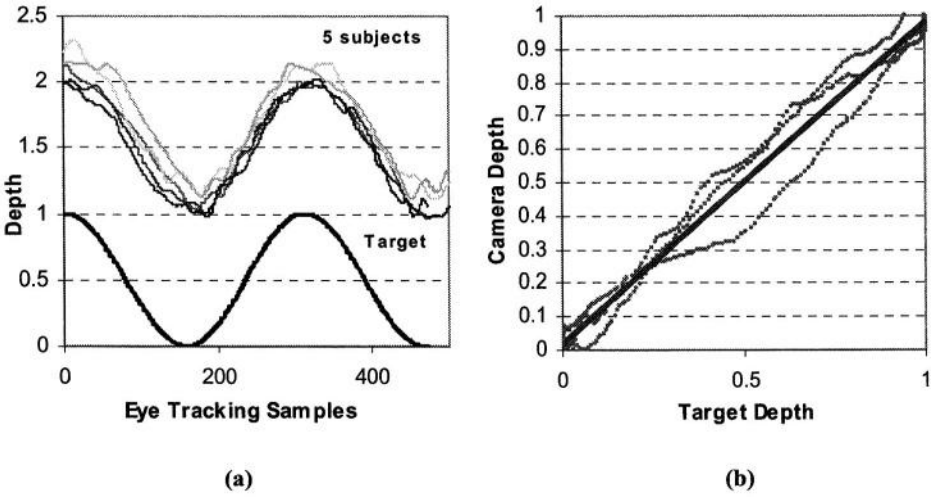


Fig. 5. (a) Gaze contingent motion compensation over a period, performed by *5* subjects. The shift of the subject lines along the depth axis corresponds to the required reference distance of the gaze-controlled camera from the target. (b) Respective linear regression plot of a subject

Table 1. Error analysis comparing the gaze contingent motion compensation performance of 5 subjects

	Regression Ratio	Bias	R^2
Subject 1	0.9052	0.1200	0.9161
Subject 2	0.9641	0.2031	0.9335
Subject 3	0.8919	0.2714	0.8751
Subject 4	0.9328	0.0876	0.9379
Subject 5	0.9692	0.0171	0.9741

4 Discussion and Conclusions

In conclusion, we have demonstrated two important features related to gaze contingent robotic control. Deploying robots around and within the human body, particularly for robotic surgery presents a number of unique and challenging problems that arise from the complex and often unpredictable environments that characterise the human anatomy. Existing master-slave based robots such as the daVinci system, which embodies the movements of trained minimal access surgeons through motion scaling and compensation, are gaining clinical significance. Under the conventional dichotomy of autonomous and manipulator technologies in robotics, intelligence of the robot is typically pre-acquired through high-level abstraction and environment

modelling. For systems that require robotic vision, this is known to create major difficulties. The ethical and legal barriers imposed on interventional surgical robots give rise to the need of a tightly integrated perceptual docking between the operator and the robot, where interaction in response to sensing is firmly under the command of the operating surgeon. The study presented here is a first step towards this goal and from our knowledge this is the first of its kind that have been demonstrated in normal subjects for both real and synthetic scenes.

It is interesting to note that for two of the subjects studied (2 and 5) for motion compensation, near perfect compensation was achieved. These particular subjects had the opportunity to spend more time performing the experiment over several sessions, suggesting experience of the system plays a certain role in the ability to stabilizing the motion of an oscillating object. It should be noted that there are a number of other issues that need to be considered for the future integration of gaze contingency to the robotic control such as dynamics of vergence [5][6] and subject/scene specific behaviour of the eye [7][8]. Other issues related to monocular preference [9], visual fatigue [10] and spatial errors that arise when portraying 3D space on a 2D window [11] will also need to be considered.

It is worth noting that for the motion compensation study it was assumed that the target oscillates along the visual axis of the camera. In a real situation, the mode of oscillation of a fixated tissue could be in any direction. Realigning the visual axis of the camera with the oscillation axis could be achieved by also taking into consideration the eye-tracking acquired oscillation components along the X-axis and Y-axis. This issue needs also further investigation.

Acknowledgements

The authors would like to thank Paramate Horkaew and Mohamed ElHelw for their valuable help.

References

1. Mon-Williams, M., Tresilian, JR, Roberts, A.: Vergence provides veridical depth perception from horizontal retinal image disparities. Exp Brain Res **133** (2000) 407-413
2. Yang, G.-Z., Dempere-Marco, L., Hu, X.-P., Rowe,A.: Visual search: psychophysical models and practical applications. Image and Vision Computing **20** (2002) 291-305
3. Bookstein, F.L.: Principal Warps: Thin Plate Splines and The Decomposition of Deformations. IEEE Trans Pattern Anal. Mach. Intell **11** (1989 June)
4. Dorf, R.C., Bishop, R.H.: Modern Control Systems, 9th Edition, Prentice Hall 2001.
5. Howard, IP, Allison, RS, Zacher, JE.: The dynamics of vertical vergence. Exp Brain Res. **116(1)** (1997 Aug) 153-9
6. Kawata, H., Ohtsuka, K.: Dynamic asymmetries in convergence eye movements under natural viewing conditions. Jpn J Ophthalmol. **45(5)** (2001 Sep-Oct) 437-44
7. Stork, S., Neggers, S.F.W., Müsseler, J.: Intentionally-evoked modulations of smooth pursuit eye movements. Human Movement Science, **21** (2002) 335-348

8. Rottach, K.G., Zivotofsky, A.Z., Das, V.E., Averbuch-Heller, L., Discenna, A.O., Poonyathalang, A., Leigh, RJ.: Comparison of horizontal, vertical and diagonal smooth pursuit eye movements in normal human subjects. Vision Res. **36(14)** (1996 Jul) 2189-95
9. van Leeuwen, A.F., Collewijn, H., Erkelens, C: Dymanics of horizontal vergence movements: interaction with horizontal and vertical saccades. Vision Res. **38(24)** (1998 Dec) 3943-54
10. Takeda, T., Hashimoto, K., Hiruma, N., Fukui, Y.: Characteristics of accommodation toward apparent depth. Vision Res. **39(12)** (1999 Jun) 2087-97
11. Wann, J.P., Rushton, S., Mon-Williams, M.: Natural problems for stereoscopic depth perception in virtual environments. Vision Res. **35(19)** (1995 Oct) 2731-6

Freehand Cocalibration of Optical and Electromagnetic Trackers for Navigated Bronchoscopy

Adrian J. Chung, Philip J. Edwards, Fani Deligianni, and Guang-Zhong Yang

Royal Society/Wolfson Foundation Medical Image Computing Laboratory,
Imperial College, London, UK
{a.chung,eddie.edwards,g.yang}@imperial.ac.uk

Abstract. Recent technical advances in electromagnetic (EM) tracking have facilitated the use of EM sensors in surgical interventions. Due to the susceptibility to distortions of the EM field when placed in close proximity to metallic objects, they require calibration in situ to maintain an acceptable degree of accuracy. In this paper, a freehand method is presented for calibrating electromagnetic position sensors by mapping the coordinate measurements to those from optical trackers. Unlike previous techniques, the proposed method allows for free movement of the calibration object, permitting C^2 continuity and interdependence between positional and angular corrections. The proposed method involves calculation of a mapping from $\{\mathbb{R}^3, SO(3)\}$ to $\{\mathbb{R}^3, SO(3)\}$ with radial basis function interpolation based on a modified distance metric. The system provides efficient distortion correction of the EM field, and is applicable to clinical situations where a rapid calibration of EM tracking is required.

1 Introduction

1.1 Background

In image-guided surgery, optical tracking has become the method of choice for coordinate measurement due to the linearity, stability and accuracy of these systems. One disadvantage associated with this technique is that line-of-sight must be maintained between the sensor markers attached to the patient or instrument. This can be inconvenient in-theatre and limits the possibilities of tracking intrustments inserted in patients. On the other hand, Electromagnetic (EM) trackers have been available for many years and offer a low cost solution for surgical navigation. A well known problem is that such systems suffer from distortions near metallic objects, which has prohibited their intraoperative use. The recent development of miniature EM trackers, however, has led to a resurgence of interest in this field. These trackers are sufficiently small (~1mm x 8mm) to be inserted into catheters or endoscopes placed inside the patient. An important use for these miniature tracking sensors is in augmented reality applications for improving diagnostic accuracy and surgical planning. Diagnosis is made more reliable, and in reduced time, when the specialist is no longer burdened with the

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 320–328, 2004.

cognitive task of fusing together multiple modes of data, such as 2D video and 3D tomographic images. Normally data fusion is achieved via the use of image registration algorithms, however most methods to date require very favourable initialisation conditions to ensure a successful registration. The advent of high precision tracking equipment compatible with in vivo usage, offers the potential of improving the robustness of current data registration techniques. Registration of 2D and 3D data is of particular importance in applications such as bronchoscopy. The ability to visualise the position of the endoscope in the CT scan has been proposed as a useful adjunct to conventional bronchoscopy [1, 2]. Further enhancement may be achieved by aligning the endoscopic images to a virtual view based on the CT scan [3] via 2D/3D registration, thus facilitating the display of virtual structures not visible in the endoscope view [4]. Reliability of the data fusion process can be improved by tracking the endoscope's location and orientation. To this end, bronchoscopes and other flexible endoscopes necessitate the use of miniaturised EM trackers in locating the viewpoint from which video is being captured. An important issue related to this approach is the presence of metal objects in the measurement volume which has a negative impact on the accuracy. Ferromagnetic interference distorts electromagnetic fields, introducing bias in measurements, hence EM tracking systems always need to be calibrated to counteract these field distortion effects.

1.2 EM Tracker Calibration

The problem of field distortion when using EM position sensors is not new. A number of methods have been proposed which correct for such distortion by using other reference position measurements. A good review of techniques is provided by Kindratenko [5], which covers virtual reality applications such as the CAVE. In these systems it is possible to remove most of the metal from the environment, thereby reducing the associated distortions. Some of the methods correct only for smooth distortions using global polynomials [6], and others provide more local interpolation of errors. In all cases, the orientation and position displacements are assumed to be dependent only on position, i.e. the errors are assumed not to change with orientation. Livingston and State, who used a look-up table with trilinear interpolation, made measurements indicating that this assumption was not valid [7]. The method they presented did not correct for orientation dependent errors as they suggested that a full 6D look-up table was impractical.

There have also been extensive clinical applications of EM calibration. Birkfellner *et al* have suggested hybrid optical and EM tracking to overcome the line-of-sight problem [8] in image-guided surgery, using Livingston and State's method. Nakada *et al* suggested a freehand method [9] using a global 4th order polynomial approximation similar to that of Ikits [6]. Wu and Taylor have stated that interdependence occurs between positional and orientation error in the case of the tracker we are using, the NDI Aurora [10]. Their solution is to provide a smooth interpolation of positional error using Bernstein polynomials and to use linear interpolation of angular error (along the arc) between different orientations of the tracker. They use an object with accurately machined grooves to

mount the sensors and the whole object can be moved by a small plastic robot to provide a regular grid of sample points.

In practice, the choice of calibration method is dependant on the application domain and for bronchoscopy tracking the calibration method needs to consider dynamic bronchial distortion due to cardiac and respiratory motion. Since the accuracy of EM trackers is reduced when tracking objects in constant motion, it is not absolutely necessary to obtain static measurements when performing the calibration. Furthermore, the method needs to be simple enough to implement in a clinical environment. Although a robotic solution shows great promise, assembling a six degree of freedom robotic arm made from EM neutral materials, in practice, introduces concerns regarding reliability, accuracy, and complexity. These considerations suggest a freehand method of calibration that involves collecting positional measurements distributed irregularly and sparsely within the measurement volume. An important challenge of any freehand method is ensuring that there is adequate coverage of the target domain which, in this case, is six dimensional. The system should be flexible and not rely on an unchanging set of static positions and orientations, hence interactive feedback is proposed so that the calibration process adapts to measurement data as they are collected. In the absence of stationary positions of measurement, sensor velocity will have to be monitored since the accuracy of the EM tracking system is reduced when the sensor is in motion. Measurements acquired during rapid motions of the sensor will have to be discounted and suitable feedback issued via the GUI. It is also important to realise that measurements will tend to be scattered irregularly over a six dimensional domain, so interpolation methods that assume rectilinear grids can no longer be applied.

2 Method

2.1 The Electromagnetic Tracker

For this study, an NDI Aurora EM tracker is used. A typical configuration of the NDI Aurora comprises of a field generator and one or more sensors. The field generator must be fixed in position and orientation in close proximity to the patient. There are two types of sensors, differentiated by the number of coils used to construct the tool. The single coil sensors are small enough for use in catheters and biopsy channel guide wires. Having only a single coil they are limited to reporting position and direction only (i.e. five degrees of freedom). Dual coil sensors must be used in larger tools since they typically contain two coils fixed at an angle to allow both position and full orientation to be reported (i.e. six degrees of freedom). The Aurora system is particularly suited to tracking flexible endoscopes during a minimally invasive intervention as shown in Figure 1. This will typically employ a single coil sensor due to size restrictions imposed by the dimensions of the biopsy channel or catheter. A 6DOF sensor is also fixed to the patient in order to correct for any patient motion during the procedure, since it is impractical to fix the patient into a rigid relationship with the field generator. The volume over which the Aurora system can effectively

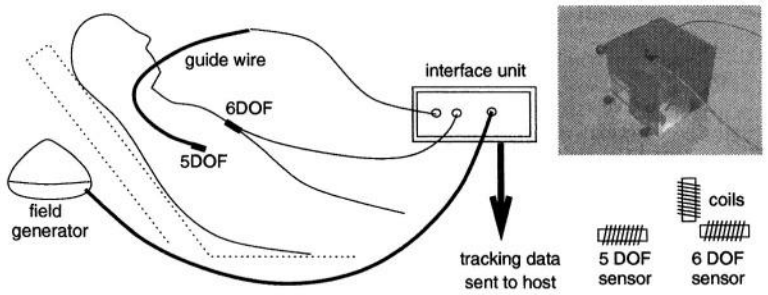


Fig. 1. A typical configuration for electromagnetic tracking systems in a clinical setting. **(inset)** Optically tracked calibration tool is shown with EM sensors attached.

track the sensor consists of a cube with 500mm sides located 50mm from the field generator. In an environment free from electromagnetic interference, the Aurora has static positional accuracy of 0.9-1.7mm and an orientation accuracy of 0.5 degrees [11].

2.2 Ground Truth

In order to correct for the error resulting from electromagnetic distortion, the absolute position and orientation must be determined by an alternative method to serve as the ground truth. An optical method was adopted that involved the tracking of infrared markers using the highly accurate stereo cameras of the NDI Polaris tracking system. In a passive configuration Polaris is capable of 0.35mm accuracy over a measurement volume of 1-2 cubic metres. To fix the two types of Aurora sensor and the passive infrared markers into a single rigid frame, a special handheld tool was machined from a 100mm wide cube of perspex shown in Figure 1 (inset).

The error correction applied to 5DOF sensors cannot, in general, be used to correct errors in 6DOF tools, because positions of individual coils are not provided by Aurora. For this reason, the calibration method must correct for errors in position, direction and orientation. To facilitate this, recesses for both 5DOF and 6DOF were machined into the surfaces of the handheld tool. Since only one face of the cube can be optically tracked, the 5DOF sensor has three places in which it can be inserted, while the 6DOF reference tracker has six places, to ensure adequate coverage of all orientations when measurements were made.

2.3 Measurement Collection

Measurements were collected by moving the perspex cube through the Aurora measurement volume while rotating the tool through a pre-planned set of orientations to ensure adequate coverage of the six dimensional space. This is difficult

to achieve in practice without computer assistance, hence a graphical interface was designed to guide the user towards all relevant areas in the measurement volume. The GUI displays a number of preset virtual targets through which the tool must pass while being rotated through a fixed set of orientations. Seldom will the user be able to hit every orientation in every target, hence the GUI can also show the specific orientations at each target that the user may have missed. The flexibility of this approach allows the system to be initialised with any arbitrary set of virtual targets. In the absence of stationary positions of measurement, sensor velocity had to be monitored and data acquired during rapid movements were discounted, thus ensuring an optimal trade off between calibration speed and measurement accuracy.

2.4 Error Correction

Corresponding measurements of position and orientation from the Aurora and Polaris tracking systems were collected. In an environment free from ferromagnetic interference there should only be a single rigid coordinate transformation relating all measurement pairs. However, in the clinical environment, the close proximity of metal introduced a non-rigid bias in the Aurora data. These discrepancies were assumed to have six degrees of freedom that were functionally dependent on both position, expressed as Cartesian coordinates $\mathbf{p} = (x, y, z)$, and orientation, expressed as a quaternion $\mathbf{q} = (q_s, q_x, q_y, q_z)$. To ensure smooth interpolation over this six dimensional space, a radial basis spline framework was extended to cater for the non-Euclidean nature of orientation space.

The error in position was modelled by the following equation:

$$\epsilon_{pos}(\mathbf{p}, \mathbf{q}) = \sum_{i=0}^{n-1} \mathbf{w}_i U(r_i(\mathbf{p}, \mathbf{q}), \sigma_i) \tag{1}$$

where $\mathbf{w}_i$ is a triple (w_{ix}, w_{iy}, w_{iz}) and the basis function, $U(r, \sigma)$, is any radially symmetric kernel. A Gaussian with a standard deviation, σ_i, was used. The distance, r_i, is measured in six-space from a finite number of control points $(\mathbf{p}_i, \mathbf{q}_i)$:

$$r_i((x, y, z), \mathbf{q})^2 = (x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2 + \alpha^2 d(\mathbf{q}, \mathbf{q}_i) \tag{2}$$

where $d(\mathbf{q}, \mathbf{q_i})$ measures distance in orientation space, and α is a scaling parameter relating positional distance with orientation distance. There are two equations for d depending on the type of tool being calibrated. For 6DOF tools, $d(\mathbf{q}, \mathbf{q}_i) = (1 - \varrho_s^2)$, where $\varrho = \mathbf{q_i}\mathbf{q}^{-1}$, calculated using quaternion multiplication. Since the orientation of 5DOF tools reported by NDI Aurora is invariant to rotation about the tool's local z-axis, for calibrating 5DOF tools $d(\mathbf{q}, \mathbf{q}_i) = (1 - \varrho_s^2 - \varrho_z^2)$

The error in orientation was corrected for by using unit quaternion splines that support simple derivatives [12]:

$$\epsilon_{ori}(\mathbf{p}, \mathbf{q}) = \omega_0 \prod_{i=1}^{n} exp(W_i(\mathbf{p}, \mathbf{q}) log(\omega_{i-1}^{-1} \omega_i)) \tag{3}$$

where ω_i are unit quaternions, $W_i(\mathbf{p},\mathbf{q}) = \sum_{j=i}^{n-1} V_j(r_j(\mathbf{p},\mathbf{q}))$, and $V_i(\mathbf{p},\mathbf{q}) = U(r_i(\mathbf{p},\mathbf{q}),\sigma_i)/\sum U_i(r_i(\mathbf{p},\mathbf{q}),\sigma_i)$. The logartihmic and exponential maps are defined by Kim *et al* [13]. For orientation error for 5DOF tools, $\boldsymbol{d}_{groundtruth} = \epsilon_{ori}\boldsymbol{d}_{aurora}\epsilon_{ori}^{-1}$ where $\boldsymbol{d}$ denotes vectors in direction space.

A clustering method was adopted to automatically select suitable values for σ_i in Eqn 2. Six-tuples of the form $(\mathbf{p}, \epsilon_{pos})$ were clustered using the k-means method which yielded positions for k control points. The orientation of each control point was set to the mean quaternion of the corresponding cluster, found by averaging rotations [14]. σ_i is the RMS 3D distance in each cluster considering the $\mathbf{p}$ components only. A similar method can be applied to per cluster quaternion distributions in choosing a value for α. This strategy has the desirable effect of automatically adapting kernel size and aspect ratio to the local error distribution in the immediate vicinity of each control point.

3 Results

Paired tracking data was collected in two different environments, an office and a hospital bronchoscopy suite. In the office environment, a 2m tall metal frame was placed at varying distances (70cm, 55cm and 40cm) from the field emitter. The frame was carefully positioned so that no part of it lay inside of the measurement volume. In the hospital environment, the Aurora was positioned behind a metal examination bed in a room containing other pieces of equipment with a high proportion of metal. In this situation, Aurora would only track sensors in 70% of the measurement volume. Since the accuracy falls when tracking fast moving objects [11], any tracking data above a threshold velocity was discarded. This left more than 600 points per dataset. For each tracking dataset, the error function ϵ was fit using half of the points, the other half being used for validation. Figure 2 summarises the results and demostrates the robustness of the technique against increasing levels of EM distortion.

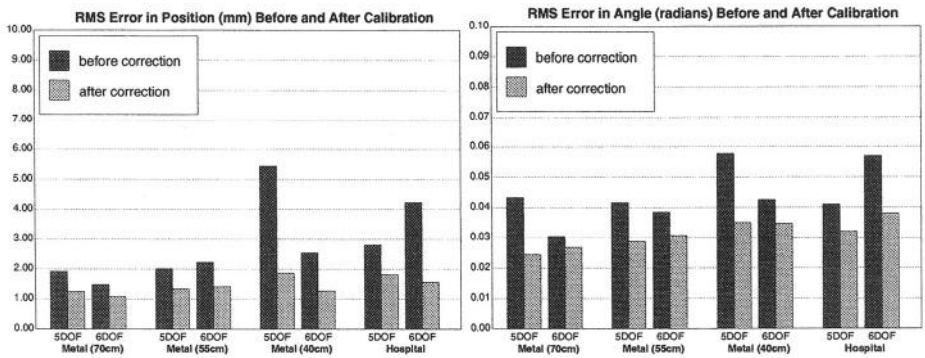


Fig. 2. RMS error in position and orientation before and after correction for 4 different datasets.

The local effect of the spline correction on positional and orientation error is illustrated in Figure 3(a). It is known that absolute error increases with distance from the field generator and this effect is exagerated by EM interference. The

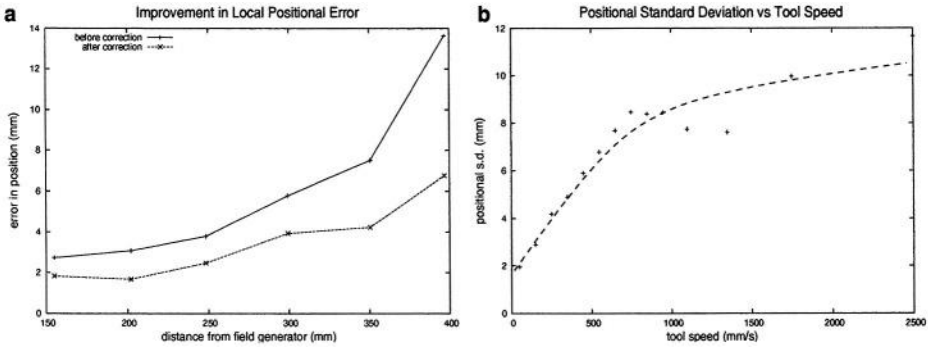


Fig. 3. (a) Local error in position before and after correction. Errors have been plotted against distance from the field generator. Electromagnetic interference was introduced via a large aluminium frame placed outside the measurement volume but within 40cm of the field generator. **(b)** Overall measurement error of the system varies with the speed at which tool is moved.

accuracy of the calibration data also depends on the speed at which the tool is moving as illustrated in Figure 3(b). The speed was estimated using tracking data reported by Polaris while Aurora tracking data was rigidly transformed to the Polaris coordinate frame for comparison with standard deviation indicating the measurement error. In summary, the technique presented can achieve an accuracy comparable to that which is achievable in an environment free from EM interference.

4 Conclusions

In this paper, a user-friendly system for calibration of an EM tracker that compensates for local field distortions has been proposed and investigated. The free-hand system has been tested in a bronchoscopy suite where metal is present in the patient couch. The methods are applicable to any situation where tracking within a patient is required in an environment that may cause EM interference. The proposed system is flexible enough to allow any configuration of 6D targets and is applicable to both 5DOF and 6DOF sensors. In contrast to previous methods, there is no need to fully populate the 6D measurement domain since the radial basis spline is designed to cope with irregular and sparsely distributed data points. Finally, by not relying on stationary sensor measurements, and exploiting the trade off between sensor velocity and accuracy, calibration time can be reduced significantly.

Acknowledgements

We would like to thank EPSRC for funding this work and Northern Digital, Inc. for their collaboration. We are also grateful to Mr David Woodman of the Medical Engineering Unit at Imperial for his help in constructing the calibration object.

References

[1] Solomon, S.B., White, P., Wiener, C.M., Orens, J.B., Wang, K.P.: Three-dimensional ct-guided bronchoscopy with a real-time electromagnetic position sensor - a comparison of two image registration methods. Chest **118** (2000) 1783–1787

[2] Schwarz, Y., Mehta, A.C., Ernst, A., Herth, F., Engel, A., Besser, D., Becker, H.D.: Electromagnetic navigation during flexible bronchoscopy. Respiration **70** (2003) 516–522

[3] Mori, K., Deguchi, D., Sugiyama, J., Suenaga, Y., Toriwaki, J., Maurer, C.R., Takabatake, H., Natori, H.: Tracking of a bronchoscope using epipolar geometry analysis and intensity-based image registration of real and virtual endoscopic images. Med. Image Anal. **6** (2002) 321–336

[4] Shahidi, R., Bax, M.R., Maurer, C.R., Johnson, J.A., Wilkinson, E.P., Wang, B., West, J.B., Citardi, M.J., Manwaring, K.H., Khadem, R.: Implementation, calibration and accuracy testing of an imageenhanced endoscopy system. IEEE Trans. Med. Imaging **21** (2002) 1524–1535

[5] Kindratenko, V.: A survey of electromagnetic position tracker calibration techniques. Virtual Reality: Research, Development and Applications **5** (2000) 169–182

[6] Ikits, M., Brederson, J.D.and Hansen, C., Hollerbach, J.: An improved calibration framework for electromagnetic tracking devices. In: Proc. IEEE Virtual Reality. (2000) 63–70

[7] Livingston, M.A., State, A.: Magnetic tracker calibration for improved augmented reality registration. Presence-Teleoper. Virtual Env. **6** (1997) 532–546

[8] Birkfellner, W., Watzinger, F., Wanschitz, F., Ewers, R., Bergmann, H.: Calibration of tracking systems in a surgical environment. IEEE Trans. Med. Imaging **17** (1998) 737–742

[9] Nakada, K., Nakamoto, M., Sato, Y., Konishi, K., Hashizume, M., Tamura, S.: A rapid method for magnetic tracker calibration using a magneto-optic hybrid tracker. In: MICCAI 2003. Lecture Notes in Computer Science (2879), Springer-Verlag (2003) 285–293

[10] Wu, X.H., Taylor, R.: A direction space interpolation technique for calibration of electromagnetic surgical navigation systems. In: MICCAI 2003. Lecture Notes in Computer Science (2879), Springer-Verlag (2003) 215–222

[11] Frantz, D.D., Wiles, A.D., Leis, S.E., Kirsch, S.R.: Accuracy assessment protocols for electromagnetic tracking systems. Phys. Med. Biol. (2003) 2241–2251

[12] Kim, M.J., Kim, M.S., Shin, S.Y.: A general construction scheme for unit quaternion curves with simple high order derivatives. Computer Graphics **29** (1995) 369–376

[13] Kim, M.J., Kim, M.S., Shin, S.Y.: A compact differential formula for the first derivative of a unit quaternion curve. The Journal of Visualization and Computer Animation **7** (1996) 43–58
[14] Curtis, W.D., Janin, A.L., Zikan, K.: A note on averaging rotations. In: Virtual Reality. Annual International Symposium '93. (1993) 377–385

3D Automatic Fiducial Marker Localization Approach for Frameless Stereotactic Neuro-surgery Navigation

Lixu Gu[1,2], Terry Peters[2]

[1] Computer Science, Shanghai Jiaotong University, Shanghai, China
gu-lx@cs.sjtu.edu.cn
[2] Robarts Research Institute, London, Ontario, Canada
tpeters@imaging.robarts.ca
www.imaging.robarts.ca/igns

Abstract. Although frameless stereotactic neuro-surgical navigation systems are widely used in many neuro-surgical centers around the world, most of the systems still require the user to define the position of fiducial markers manually from patient scans, a procedure that is tedious, time consuming and often inaccurate. Some researchers have addressed this problem, but they acknowledge that their 2D image processing approach has limitation. We propose a new automatic approach for 3D localization of the fiducial markers, which provides higher 3D localization accuracy, and is independent of the geometry of the marker. Our approach includes three steps. First, sets of 3D morphological operations are employed to extract the candidate fiducial markers as the "seeds". Then a "conditional dilation" technique is employed to reconstruct the regions of fiducials from the "seeds" which are sifted by several knowledge-based rules. Lastly, the intensity-weighted centroid of each extracted fiducial region is calculated as our final fiducial position. The approach is validated by simulated datasets and a CT phantom scan where the average *Fiducial Localization Error* (FLE) is 0.37mm and 0.31mm, respectively.

1 Introduction

When performing minimally invasive surgical interventions, the surgeon's direct view is restricted. Recent progress in computerized imaging techniques has provided pre- and intra-operative images which are exploited to obtain information about the interior of the body. This has increased the use of minimally invasive techniques in general and in brain surgery [1] in particular. Image-guide procedures [2,3] are being employed with increasing frequency in the operating room.

A fundamental requirement for image-guided surgery is that the preoperative images be precisely registered with the patient. Many stereotactic systems employ a frame to satisfy the need for accurate co-registration and probe guidance. However, this approach is often limited since the presence of the frame can be a physical constraint during surgery. The use of computer-based tracking systems using skin mounted or implantable markers provide us with means to address this limitation. These so-called "frameless" stereotactic systems provide the surgeon with naviga-

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 329-336, 2004.

tional information, relating the location of instruments in the operative field to preoperative image data without the use of a frame.

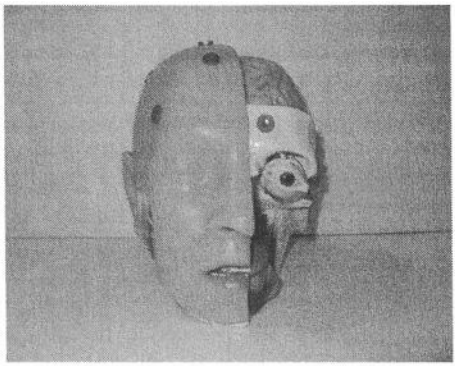


Fig. 1. Fiducial markers attached to a head phantom

There are four different techniques used in co-registration of current frameless stereotactic systems [4]: Point-based methods; Edge methods; Moment methods and "Similarity criterion optimization" methods. Point-based methods using fiducial markers are considered to be quick and reliable [5], and represent the most commonly used approaches.

Several kinds of fiducial markers are employed in current image-guided surgery systems. Some of them are rigidly implanted in the skull [6], but are both time-consuming to apply (requires a separate surgical procedure) and painful for the patient. An alternative to implanted marker is the fiducial that is attached to the skin on the patient's head. An example of this kind of marker (Aesculap -- Tuttlingen, Germany) attached on a phantom head is shown in Fig.1.

We employ the term *Fiducial localization* as the determination of the centroid of a fiducial marker in the acquired image. Wang et al [7] described a method to localize the implanted fiducials using a knowledge-based technique, which is limited to 2D image processing and fiducial markers of a particular geometry. In this paper, we propose an automatic fiducial localization approach using a set of fully 3D morphological techniques. The approach is validated using simulated datasets and a phantom CT scan. Furthermore, our method is not restricted to a particular geometry of fiducial marker.

The rest of the paper is organized as follows: in section 2, after a brief review of mathematical morphology, we describe a fiducial marker detection algorithm using 3D morphological segmentation techniques. In section 3, an intensity-weighted centroid is introduced to determine the fiducial positions, and experiments using both simulated datasets and a CT phantom scan are implemented to validate the proposed approach in section 4.

2 Morphological Treatment for the Detection of Fiducial Markers

2.1 Morphological Operations

Mathematical morphology is a powerful technique for the quantitative analysis of geometrical structures. It consists of a broad and coherent collection of theoretical concepts, nonlinear signal operators, and algorithms aimed at extracting objects from images.

We define a 3D image f as a subset of the 3D Euclidean space ($f \in \mathbf{R}^3$), and a 3D structuring element $k \in \mathbf{R}^3$. The four basic operations can be defined as follows:

$$\text{Dilation:} \quad f \oplus k = \bigcup_{b \in k} (\{a + b \mid a \in f\}) \tag{1}$$

$$\text{Erosion:} \quad f \ominus k = \bigcap_{b \in k} (\{a - b \mid a \in f\}) \tag{2}$$

$$\text{Opening:} \quad f \circ k = (f \oplus k) \ominus k \tag{3}$$

$$\text{Closing:} \quad f \bullet k = (f \ominus k) \oplus k \tag{4}$$

Many other morphological algorithms are derived from these four operations. *Top-hat* transformation (shown in Fig.2) and *Conditional Dilation* (*C-Dilation*) are two typical algorithms employed in our approach for feature extraction and region reconstruction, respectively. They are defined as:

$$\text{Top-hat:} \quad T(f,k) = f - (f \circ_{gray} k) \tag{5}$$

$$\text{C-Dilation:} \quad B_i = (B_{i-1} \oplus k) \cap | f |_G \qquad (B_i \in R^3, i = 1,2,\ldots). \tag{6}$$

Here, $\circ_{gray}$ denotes an Opening operation on a grayscale image, and $| f |_G$, the *mask* of the operation, is the result of a threshold operation using gray level *G*. The iteration in (6) is repeated until there is no change between B_{i-1} and B_i.

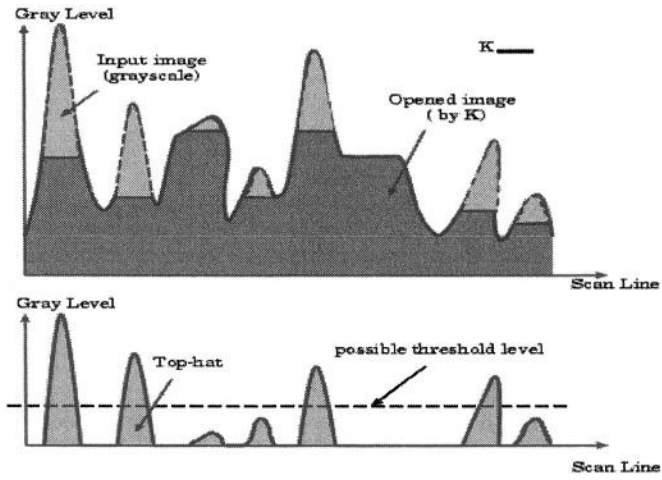


Fig. 2. Top-hat transformation in one dimension. Where dark gray area in upper figure stands for opened image, and the difference between the source and opened images is depicted in the lower figure as the TT result.

3 Fiducial Marker Detection

Our fiducial marker detection algorithm is based on the 3D morphological segmentation techniques outlined above. Processing is divided into three steps: candidate fidu-

cial marker region detection, fiducial marker seed reconfirmation and fiducial marker region recovery.

The first step is designed to detect fiducial marker regions from an entire 3D data set using a top-hat transformation (TT). Since the fiducial markers are designed to present a high intensity level in the input data set and have a standard dimension, the TT can effectively detect them using a 3D spherical structuring element with a size just greater than the largest dimension of the fiducial markers. This detection approach is independent of the shape information of the fiducials. A concept description of the TT function is shown in Fig.2 and defined by equation (5). After we calculate the $T(f,k)$, it is threholded into a binary image $|T|_{G1}$ to distinguish the fiducials from other parts of the image. The intensity level G_1 is determined by the histogram of the TT result $T(f,k)$. Then a binary Opening operation using a spherical structuring element with a diameter of 3 is employed to reduce the remaining noise. This procedure provides us with our candidate fiducial "seeds". The entire processing can be interpreted as a 3D sphere employed to search the 3D input data space to find the objects within a specific high intensity level and with a size smaller than the sphere.

To avoid missing detected fiducial seeds, we confirm every fiducial candidate in the second step, using the following criteria:

1. The distance between every two fiducial seeds should be larger than a constant D. Since fiducial markers are required to be distributed as evenly as possible over an approximately spherical surface, seeds with spacing smaller than D are considered as error candidates which need to be further identified by the next criterion.
2. The intensity of the fiducial seeds in the source image should be in the same range. Since the fiducial markers are made of the same material, a fiducial candidate with an intensity value out of the reasonable range is considered to be an artifact. We note that for MR images, the image should be corrected for rf-inhomogeneities, to ensure that this condition is met.

Such erroneously detected seeds R_{error} are discarded during this step and the final entries E are stored:

$$E = |T(f,k)|_{G1} \circ k_{sphere\_3} - R_{error} \quad , \tag{7}$$

which are pushed into the next step to accurately define the fiducial regions.

During the threshold operation in TT and the noise reduction processing in equation (7), parts of the fiducial contours are destroyed. A morphological "C-Dilation" algorithm is then employed to recover these lost regions. The concept of the *c-dilation* function is defined as equation (6), where the *mask* is defined as $|f|_{G0}$ when the starting *Marker* $M = B_0 = E$ in the iteration. The threshold level G_0 employed here is determined by a histogram analysis of the source image. For CT, G_0 is set to a value just higher than that of bone value resulting a *mask* containing the complete fiducial regions. When the condition of $B_i = B_{i-1}$ is satisfied, the iteration stops automatically, and the final fiducial regions are obtained accurately.

4 Localization Approach

The only assumption we make about the fiducial markers is that they incorporate a spherical depression, matched to a probe tip, which is precisely located at the centroid of the marker. In this manner, accurate identification of the fiducial centroid will at the same time identify the position located by a probe.

Intensity-weighted centroids of each extracted fiducial marker region are used as our fiducial localization results. Intensity weighting implies that the voxel coordinates of the fiducial components are weighted by their intensity. First, the final extracted binary fiducial regions are converted to gray scale by a voxel-based multiplication with the source image. Then the centroid coordinates (X, Y, Z) of the intensity weighted fiducial regions are calculated separately:

$$X = \frac{\sum_{i=0}^{i\max} x_i A(x_i)}{\sum_{i=0}^{i\max} A(x_i)}; \qquad A(x_i) = \sum_{j=0}^{j\max}\sum_{k=0}^{k\max} I(x_i, y_j, z_k) \tag{8}$$

$$Y = \frac{\sum_{j=0}^{j\max} y_j A(y_j)}{\sum_{j=0}^{j\max} A(y_j)}; \qquad A(y_j) = \sum_{i=0}^{i\max}\sum_{k=0}^{k\max} I(x_i, y_j, z_k) \tag{9}$$

$$Z = \frac{\sum_{k=0}^{k\max} z_k A(z_k)}{\sum_{k=0}^{k\max} A(z_k)}; \qquad A(z_k) = \sum_{i=0}^{i\max}\sum_{j=0}^{j\max} I(x_i, y_j, z_k) \tag{10}$$

Here, $I(x_i, y_j, z_k)$ is the intensity value of a voxel in the fiducial region, and *imax, jmax* and *kmax* are the numbers of voxels in the X, Y, Z directions in a fiducial region respectively.

5 Validation

5.1 Modeling Study

To validate the proposed approach, we created a series of 3D volume datasets with ten model fiducial markers in each, for use in a simulation experiment. The size of each dataset was 512×512×120 with *1mm* spacing between pixel centers. We employ a cylinder-shaped object with radius r = *6mm* and height h = *4mm* as the fiducial marker, which is the exact dimension of a typical real fiducial marker shown in Fig.1. The whole modeling processing is described as follows.

After the 3D volume is created, ten fiducial centers, randomly selected in 3D space, are recorded as the target fiducial positions. The ten fiducials are automatically

brought into the data volume one by one in a random manner according to the following algorithm:

1. Create a 20 times enlarged cylinder-shaped fiducial marker in a 300×300 ×300 temporary volume region with a size of $r = 120$ and height $h = 80$, where the center of the fiducial is exactly at the center position of the temporary region. This enlargement is designed to model the fiducials more precisely.
2. Rotate the fiducial marker randomly in three directions.
3. Shrink the whole temporary region back to the original dimension with *1mm* pixel distance. We assign the average value of every 20 voxels, to the new corresponding voxel in the shrunken region where the model fiducial marker is finally created. This step simulates the partial volume phenomenon that occurs in practice.
4. Place the created fiducial back to the 3D volume dataset at the position of the corresponding fiducial center.
5. Repeat for all markers in the volume.

In the second step, we blur the created dataset by a *normal distribution* convolution kernel:

$$g(x) = \frac{1}{\sqrt{2\pi}\delta} \exp[-\frac{1}{2}(\frac{x-m}{\delta})^2], \tag{11}$$

where, we set $m = 0$, and δ range from 0.5 to 2.0 to evaluate the effect of different image resolutions. The convolution which is accomplished in 3D space, is defined as follows:

$$f[m,n,o] \otimes g[i,j,k] = \sum_{a=0}^{i-1} \sum_{b=0}^{j-1} \sum_{c=0}^{k-1} f[m-a, n-a, o-c] g[a,b,c] \tag{12}$$

5.2 Phantom Study

A CT scan using the head phantom shown in Fig.1 was employed to further validate our proposed approach, and in particular to demonstrate that our thresholding and morphological operations could robustly separate the markers from underlying structures. This image contains 45 slices with 3mm thick; each slice contains 512×512 pixels of 0.48mm spacing; 8 cylinder-shaped fiducials are randomly attached to the phantom surface, where a concave pivot is located at fiducials' center for pointer rest. Three enlarged cross section views of an example fudicial (#2) are shown in Fig.3, where the highlighted dots labeled with number "2" indicate the automatically detected fiducial position using our proposed approach. Detected fiducial markers with labels can also be identified from the 3D volume viewport.

5.3 Experimental Results

Experiments using a series of model datasets and a CT phantom scan were implemented to evaluate our proposed algorithm. Model datasets include an *in-plane* (without rotation) *fiducial dataset* **(IPFD),** a *dataset without blur* **(DWOB),** four

datasets with blur ranged from $\delta = 0.5$ to 2.0 **(DWB0.5 ~ DWB2.0),** and a *full dataset with rotation and blur* ($\delta = 1.0$) **(FDRB).** *Fiducial Localization Error* (**FLE**) [6] is employed here to measure the accuracy of the proposed fiducial localization approach, which is defined as the Euclidean distance (in mm) between *detected fiducial position* (DFP) and the randomly selected *target fiducial position* (TFP). For the phantom scan, we employ manually defined fiducial position as the TFP when the automatically detected fiducial position as the DFP. The results are shown in Tab.1, where FM1 ~ FM10 refer to the 10 fiducial markers in the experimental datasets.

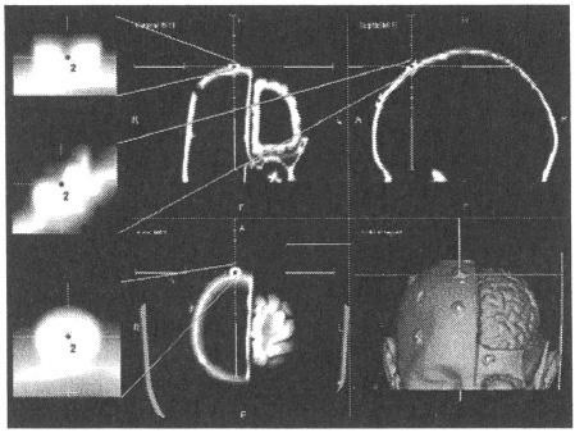


Fig. 3. Fiducials in a CT phantom scan with their TFP

Table 1. Fiducial localization errors (FLE) (in mm)

FLE (MM)	FM_1	FM_2	FM-3	FM_4	FM_5	FM_6	FM_7	FM_8	FM_9	FM_10	Average
IPFD	0.29	0.16	0.24	0.33	0.23	0.39	0.29	0.35	0.22	0.26	0.27
DWOB	0.42	0.37	0.49	0.34	0.39	0.19	0.46	0.44	0.36	0.38	0.38
DWB0.5	0.32	0.39	0.45	0.4	0.53	0.19	0.41	0.42	0.47	0.32	0.39
DWB1.0	0.41	0.23	0.53	0.42	0.37	0.48	0.27	0.43	0.35	0.42	0.39
DWB1.5	0.37	0.49	0.33	0.21	0.54	0.38	0.5	0.44	0.32	0.35	0.39
DWD2.0	0.35	0.41	0.39	0.46	0.48	0.35	0.42	0.43	0.23	0.46	0.40
FDRB	0.43	0.29	0.39	0.35	0.34	0.44	0.52	0.35	0.38	0.41	0.39
Average	0.37	0.33	0.40	0.36	0.41	0.35	0.41	0.41	0.33	0.37	0.37
Phantom	0.3	0.2	0.4	0.3	0.4	0.2	0.3	0.4	--	--	0.31

6 Discussion

The 3D localization technique for fiducial markers described in this paper is based on full 3D morphological segmentation algorithms, which reduces the dependence on the shape and the orientation of the 3D fiducial markers. In contrast, the 2D segmentation algorithm described in [7] depends on the orientation of the marker, and the orientation has an effect on the shape of the 2D cross section. Moreover, it may result in a marker cross-section not being identified.

We are undertaking further experiments using both the phantom and patient CT and MRI scan to validate the proposed approach, where a variety of fiducial marker geometries will be tested, and validated against frame-based localization approaches.

7 Conclusion

An automated fiducial marker localization algorithm is proposed for image registration in frameless stereotactic neuro-surgery navigation. The approach is designed to work with different types of fiducials using multi-modality images. The algorithm is validated by a series of simulated datasets and a CT phantom scan showing FLE is in the order of 0.37mm and 0.31mm, respectively.

References

1. A. D.Salles, R.Lufkin (eds), *"Minimally Invasive Therapy of the Brain"*, Thieme, New York, Stuttgart, 1997.
2. F.A.Jolesz, *"Image-guided Procedures and the Operating Room ofthe Future"*, Radiology 204:601-612,1997.
3. K.W.Finnis, Y.P.Starreveld, A.G.Parrent, T.M.Peters, "A 3-Dimensional Database of Deep Brain Functional Anatomy, and Its Application to Image-Guided Neurosurgery", In MICCAI 2000,1-8, Oct. 2000.
4. C.R.Maurer, J.M.Fitzpatrick, *"A Review of Medical Image Registration"*, Interactively active Image-Guided Neurosurgery, a book edited by R.J.Maciunas, Park Ridge, IL: American Association of Neurological Surgeons, 17-44, 1993.
5. J.West, J.M.Fitzpatrick, M.Y.Wang, B.M.Dawant, C.R.Maurer et al, *"Comparison and evaluation of retrospective intermodality brain image registration techniques"*, Journal of Computer Assisted Tomography, 21 (4): 554-566, 1997.
6. C.R.Maurer, J.M.Fitzpatrick, M.Y.Wang, R.L.Galloway, R.J.Maciunas, G.S. Allen, *"Registration of head volume images using implantable fiducial markers"*, IEEE Transactions on Medical Imaging, 16:447-462, 1997.
7. M.Y.Wang, C.R.Maurer, J.M.Fitzpatrick, R.J.Maciunas, *"An automatic Technique for Finding and Localizing Externally Attached Markers in CT and MR Volume Images of the Head"*, IEEE Transactions on Biomedical Engineering, vol. 43, No.6: 627-637, June 1996.

Contact Modelling Based on Displacement Field Redistribution for Surgical Simulation

Bryan Lee[1,2], Dan Popescu[1], and Sébastien Ourselin[1]

[1] CSIRO ICT Centre, BioMedIA Lab
Cnr Vimiera & Pembroke Rds, Marsfield NSW 2122, Australia
http://www.ict.csiro.au/BioMedIA/
{BryanC.Lee, Dan.Popescu, Sebastien.Ourselin}@csiro.au
[2] School of Electrical and Information Engineering
University of Sydney, Bldg J03, Sydney NSW 2006, Australia
http://www.ee.usyd.edu.au/

Abstract. We present a novel method of contact modelling for discrete deformable models. Our algorithm is used to simulate contact between rigid surgical tools of arbitrary shape and deformable virtual organs bounded by triangular mesh surfaces. It uses a divide and conquer strategy to redistribute an arbitrary field of displacements on organ surfaces into an equivalent field of displacements at the nodes. The computational complexity depends on the size of the touch field, but not on the size of the mesh representing the organ. Our algorithm results in accurate modelling and can be used in real-time applications requiring haptic feedback.

1 Introduction

Modelling of the interaction between virtual organs and virtual surgical tools is a key aspect in the design of a surgical simulation system. See Basdogan *et al.* [1] for a recent review on the major issues of this complex topic, including interaction with soft tissue. In general, this interaction can be described by the three major components shown schematically in Fig. 1: *collision detection, contact modelling* and *collision response.* Collision detection identifies the objects of a virtual scene that are in contact. Contact modelling determines the shape of the area of contact and possible field of interpenetrations, in the case of deformable objects. Collision response determines how the objects react to contact, by moving, deforming and/or providing force feedback.

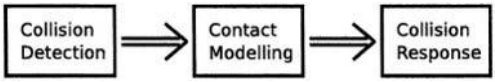


Fig. 1. *The three major components of modelling interaction.*

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 337–345, 2004.

While collision detection and collision response have been studied extensively, contact modelling for deformable models has received relatively little attention in the literature.

Hansen *et al.* [4] address the problem of contact modelling for organs and tools represented as discrete meshes, by inserting extra vertices to accurately model the deformation. Adding vertices means the models used by the collision response algorithm have to be restructured. In the case of a finite element method (FEM) collision response algorithm, the model matrices are altered and have to be re-calculated, causing a significant performance degradation. The authors therefore propose an algorithm that partially moves vertices close to the tool, based on a simple linear relationship. This method provides a set of displacements on nodes that can be tuned for increased realism. The displacements result in a smooth deformation calculated in a fast and efficient manner. However, tuning of the algorithm is not trivial and is dependent on the nature of the organ being modelled. Furthermore, the algorithm operates on nodes and does not consider contacts between the tool and the surface triangles. This results in inaccurate approximations in the modelling of the contact, especially if there are only surface triangles, and no nodes, involved in the collision. It also does not consider arbitrarily shaped or multiple tools.

Picinbono *et al.* [5] note the limitations of a common technique for contact modelling, referred to as the penalty method in the literature. This method applies a force, proportional to the penetration depth, to all triangles in collision with the tool. The authors, however, state that this method does not respect the geometry of the contact and does not guarantee that the collided faces will follow exactly the displacement of the tool, and therefore choose a geometric approach, rather than a physical one. A projection plane is defined, based on the contact area, the direction of the tool and the average normal of all collided faces. The vertices are then projected orthogonally on the normal to the projection plane. This method works well for contact made using a tool of simple shape. It also utilises information about the surface triangles rather than just the nodes, to provide a more realistic result, and ensure no interpenetrations will occur after displacement of the nodes. It is however not general enough to consider contacts made using a tool of complex shape. Furthermore, all vertices are projected in the same direction, leading to approximations that reduce accuracy.

A displacement-velocity correction method of contact modelling has been developed by Raghupathi *et al.* [7]. This method follows on from a collision detection method that focuses on virtual intestinal surgery. Due to the nature of intestines, it is designed to cope with multiple collisions and self-collision. Both the collision detection and response algorithms work with pairs of segments. This is however limiting for contact modelling and the paper states that updating the position and velocity of one edge may create or cancel other collisions. A suggestion made by the authors, to overcome this deficiency, is to repeat the collision check for all pairs, with a limit on the number of iterations to avoid infinite loops. While effective for intestinal models, the method cannot be easily generalised for models of arbitrary shapes.

Berkley *et al.* [2] use a constraints approach that allows contacts with any surface triangle, and redistributes the displacements on these triangles to equivalent displacements on all nodes in the model. While this method results in an accurate model of the contact, redistribution of all the nodes in the model is computationally too demanding, and renders it difficult to use in real-time applications.

In this paper, we propose a novel and general method of contact modelling, which can redistribute, in real-time, arbitrary contact on discrete models, in a physically accurate manner. Section 2 details the motivation for our technique and describes the mathematical model. We validate our method in Section 3, in conjunction with an FEM. We include illustrations of contact modelling results and a performance evaluation of the real-time capabilities of our technique, which demonstrates that it can be applied to surgical simulation with haptic feedback. We present our conclusions in Section 4.

2 Contact Modelling

Typically, data structures representing virtual organs will have discrete representations. In order to provide an accurate collision response, most discrete deformable models require that the input interaction be defined at the exact locations of those discrete elements. For example, in order to accurately model a virtual elastic organ as a discrete mesh, and deform it based on an FEM, a set of vertex displacements at the mesh nodes is needed [3]. The accuracy of modelling could be compromised if, say, the surgical tool is too thin for the resolution of the virtual organ, as it could penetrate the surface of the organ without actually making contact with any of the mesh nodes.

To overcome such problems, we propose a novel and general algorithm of contact modelling for deformable objects, which redistributes an arbitrary field of displacements, located anywhere on the surface of an organ, into an equivalent field of displacements located at the discrete nodes. We assume that the organ's surface can be described by a triangular mesh and that contact with a rigid surgical tool can be described as a field of displacements imposed on that mesh. This paradigm of contact modelling is universal because, in general, neither the surface of the organ nor the surface of a generic surgical tool can be described by analytical equations. There are no restrictions on the displacement vector field; therefore, our method can effectively model contacts between virtual organs and virtual surgical tools of arbitrary shape. Our method is in the same spirit as the method suggested by Berkley *et al.* [2], but uses a divide and conquer strategy that keeps the contact local. Therefore, the size of the computation is kept low and our method becomes suitable even for very demanding real-time applications with haptic feedback (1 kHz or more). The diagram of Fig. 2 details the contact modelling function.

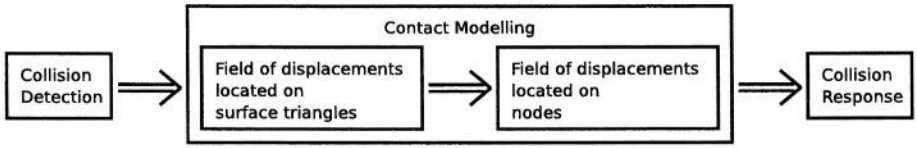


Fig. 2. *The three major components of modelling interaction, with detail of the Contact Modelling component.*

2.1 Mathematical Model

The construction is easier to understand if we firstly analyse the simple case pictured in Fig. 3. Suppose that the triangle with nodes labelled 1, 2 and 3 is touched such that the point P, having baricentric coordinates b_1, b_2 and b_3, is displaced by a vector $\mathbf{v}$. We are interested in determining what displacements $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{u}_3$ at the triangle nodes would give such a displacement of P.

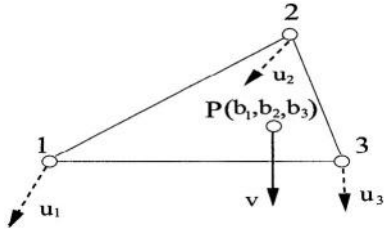


Fig. 3. *Displacing point P by vector* $\mathbf{v}$.

If we use a model of linear interpolation, then:

$$b_1\mathbf{u}_1 + b_2\mathbf{u}_2 + b_3\mathbf{u}_3 = \mathbf{v} \tag{1}$$

With no additional restriction, Eq. 1 is overdetermined if one tries to solve for $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{u}_3$ in terms of $\mathbf{v}$, with some solutions corresponding to huge vertex displacements. The solution becomes unique and "physically natural" if we adopt the displacement of "minimal energy". That is the solution of Eq. 1 which minimises $||\mathbf{u}_1||^2 + ||\mathbf{u}_2||^2 + ||\mathbf{u}_3||^2$. To solve this conditional extremum problem, we have to minimise the function:

$$F(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \boldsymbol{\lambda}) = ||\mathbf{u}_1||^2 + ||\mathbf{u}_2||^2 + ||\mathbf{u}_3||^2 + <\boldsymbol{\lambda}, \mathbf{r}> \tag{2}$$

where $\boldsymbol{\lambda} = (\lambda_x, \lambda_y, \lambda_z)^T$ is a vector of Lagrange multipliers, $<,>$ denotes scalar product and $\mathbf{r} = b_1\mathbf{u}_1 + b_2\mathbf{u}_2 + b_3\mathbf{u}_3 - \mathbf{v}$ is a "restriction" vector. Equating to zero the partial derivatives of F we get the equations:

$$\frac{\partial F}{\partial \mathbf{u}_i} = 2\mathbf{u}_i + b_i\boldsymbol{\lambda} = 0 \qquad (\mathrm{i} = 1,2,3) \tag{3}$$

$$\frac{\partial F}{\partial \boldsymbol{\lambda}} = \mathbf{r} = b_1\mathbf{u}_1 + b_2\mathbf{u}_2 + b_3\mathbf{u}_3 - \mathbf{v} = 0 \tag{4}$$

By solving the linear system we get:

$$\mathbf{u}_i = \frac{b_i\mathbf{v}}{b_1^2 + b_2^2 + b_3^2} \qquad (\mathrm{i} = 1,2,3) \tag{5}$$

which we call the L^2 *baricentric extrapolation* of $\mathbf{v}$.

We now treat the general case, where several triangles of the mesh are touched, by generalising the L^2 baricentric extrapolation. Suppose n mesh nodes belong to these touched triangles, and that we want to move l points inside these triangles – which will impose l "restrictions" similar to Eq. 1. Restriction vector $\mathbf{r}_j$ will take the form:

$$\mathbf{r}_j = \sum_{i=1}^{n} b_{ij}\mathbf{u}_i - \mathbf{v}_j = 0 \tag{6}$$

where b_{ij} is the baricentric coordinate of node i in restriction j. (For each restriction equation, only 3 coefficients b_{ij} can be non-zero.) Let us remark that if a touched triangle has a fixed node, say node s, we only need to add the restriction $\mathbf{u}_s = 0$, to maintain consistency.

We find the displacements $\mathbf{u}_i$ at nodes by minimising the function:

$$F = \sum_{i=1}^{n} ||\mathbf{u}_i||^2 + \sum_{j=1}^{l} < \boldsymbol{\lambda}_j \, , \mathbf{r}_j > \tag{7}$$

where $\boldsymbol{\lambda}_j$ is the Lagrange multiplier 3-vector corresponding to restriction vector $\mathbf{r}_j$, $(j = 1, ..., l)$. By taking the partial derivatives we get:

$$\frac{\partial F}{\partial \mathbf{u}_i} = 2\mathbf{u}_i + \sum_{k=1}^{l} b_{ik}\boldsymbol{\lambda}_k = 0 \qquad (\mathrm{i} = 1, \ldots, \mathrm{n}) \tag{8}$$

$$\frac{\partial F}{\partial \boldsymbol{\lambda}_j} = \mathbf{r}_j = \sum_{k=1}^{n} b_{kj}\mathbf{u}_k - \mathbf{v}_j = 0 \qquad (\mathrm{j} = 1, \ldots, \mathrm{l}) \tag{9}$$

Denote by B the $n \times l$ sparse matrix $B = (b_{ij})$. Eqs. 9 can be expressed compactly as:

$$(B^T U) = \mathbf{V} \tag{10}$$

and by substitution from Eqs. 9 one gets:

$$(B^T B)\boldsymbol{\Lambda} = -2\mathbf{V} \tag{11}$$

where $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1 \, , \boldsymbol{\lambda}_2 \, , ..., \boldsymbol{\lambda}_l \,)^T$ is the vector of Lagrange multiplier 3-vectors and $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_l)^T$. It follows that:

$$\boldsymbol{\Lambda} = -\frac{1}{2}(B^T B)^{-1}\mathbf{V} \tag{12}$$

and then the node displacements $\mathbf{u}_i$ are easily calculated from Eq. 8. The outlined procedure can be recognised as a Moore-Penrose pseudoinversion of Eq. 10 for the particular case when $B^T B$ is non-singular. This will be the case with most well-conditioned input displacement fields. When this does not happen, due to small inconsistencies in the input displacement field, $\mathbf{U}$ can still be obtained from Eq. 10 by using a general pseudoinverse procedure, based on singular value decomposition.

3 Validation

A contact modelling system has been implemented, based on the mathematical model described previously. To clearly illustrate the results of our contact modelling algorithm, the displacements will be applied to a series of tetrahedral meshes, of various resolutions, having the simple geometrical shape of a cube. The simplest cube is constructed from 8 nodes, 12 surface triangles and 6 tetrahedra. The higher resolution cube has 1331 nodes, 1200 surface triangles and 6000 tetrahedra.

3.1 Qualitative Results

For illustration purposes, our general method of contact modelling has been coupled to a collision response module implemented by an FEM technique acting on tetrahedral meshes and modelling linear elastic response. This is a fast method allowing haptic and visual interaction, based on a real-time inversion of a local contact matrix. Further details of this technique are described in Popescu *et al.* [6]. However, we emphasise that our contact modelling method is general and not restricted for use with only one particular collision response module.

Fig. 4 illustrates the results of our implementation on the two cubes of different resolutions. Notice how the deformations rigorously follow the imposed displacements after the contact field redistribution.

Arrows used in the figure represent the vectors of displacements. The thick arrows correspond to displacements applied to points on surface triangles and the thin arrows correspond to displacements applied to mesh nodes. In our examples, the anchored nodes are indicated by small spheres. Each row pictures a single scenario of the deformation. The first two scenarios are simple, to clearly illustrate the effect of our contact modelling method. The first one shows two vector displacements on the same surface triangle and the second one shows two vector displacements on adjacent surface triangles. Note that the surface triangles' contact points have been displaced to the tip of the arrows as desired.

The third and fourth scenarios use the higher resolution cube to illustrate our algorithm applied to larger tetrahedral meshes. Although our algorithm is independent of the size of the mesh, for a more visually reflective result, this cube is not created to be as large as those models that are typically used in surgical simulation. The third scenario illustrates a torque effect from two displacement vectors and the fourth scenario shows multiple displacements to illustrate the capability of our contact modelling algorithm.

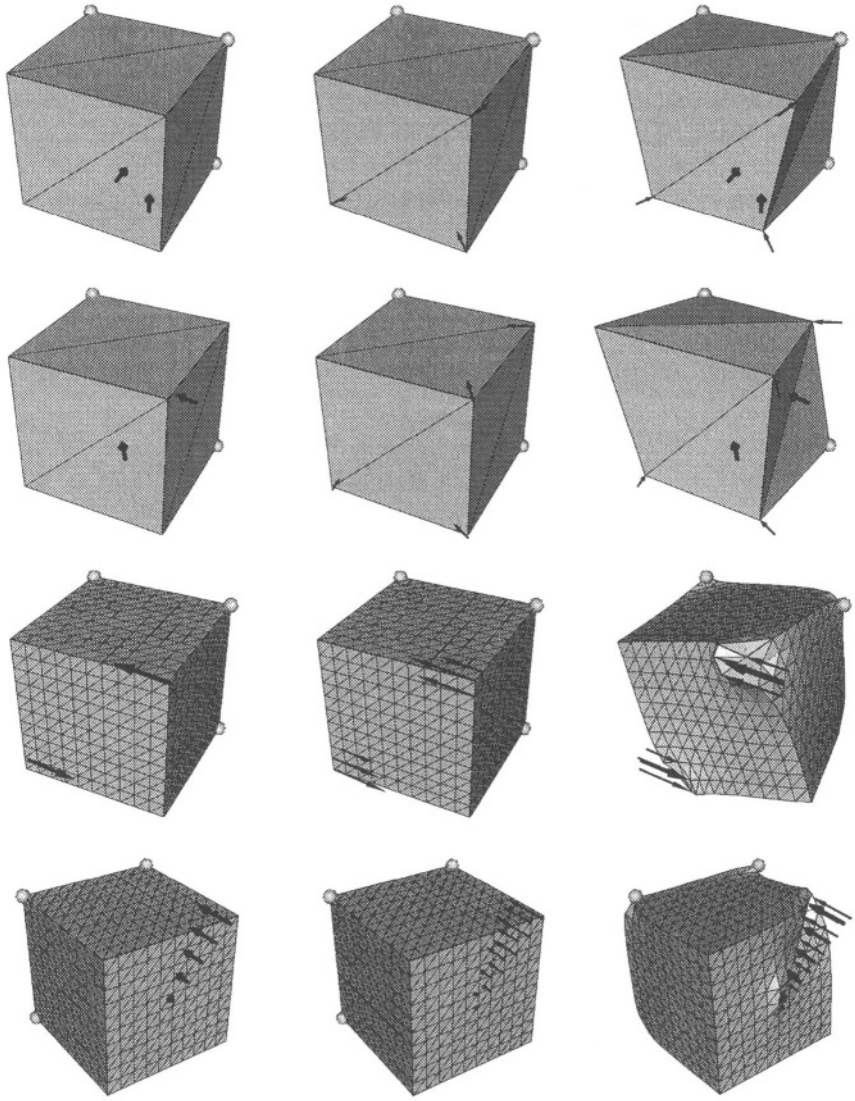


Fig. 4. *Illustration of our general method of contact modelling, with anchored nodes marked by a small sphere. Left column: the object before deformation, with thick arrows indicating the vectors to be applied as displacements on the surface triangles. The base of the arrow is located at the surface point to be displaced, and the tip points to the position where it should be moved. Middle column: the object is presented again before deformation, with thin arrows indicating the equivalent displacements of nodes, as calculated by our algorithm. Right column: the object is deformed, using the outputs of our contact modelling algorithm as inputs to our collision response module, implemented by an FEM. Both sets of arrows from the previous two illustrations are included as well.*

3.2 Performance Evaluation

The performance of our contact modelling algorithm has been measured on a 2.4 GHz Pentium 4 and a 3.0 GHz Xeon. The results are shown in Fig. 5. These results were generated by running random collisions on our higher resolution cube. The experiments were carried out with between 1 and 60 arbitrary displacements on surface triangles. For each of these experiments, 100 different seeds were used for the randomization, and each of these averaged over 1000 trials.

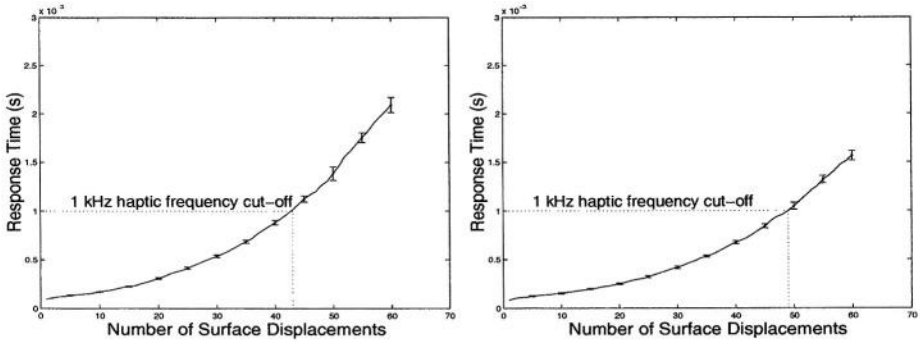


Fig. 5. *Graphs of the performance of our contact modelling algorithm on different machines. Each graph shows the response times, in seconds, compared to the number of surface displacement vectors applied. Left: Computations performed on a 2.4 GHz Pentium 4. Right: Computations performed on a 3.0 GHz Xeon.*

An analysis of our algorithm reveals the matrix inversion of Eq. 12 to be the most significant computational cost, with a standard matrix inversion based on Gaussian elimination having $o(n^3)$ computational complexity. The size n of our matrix to be inverted is proportional to the number of restriction vectors. Indeed, the graphs also have a cubic nature, which is consistent with our algorithm requiring the matrix inversion.

As haptic applications require a refresh frequency of 1 kHz, we can identify the limits of our contact modelling algorithm. On the 2.4 GHz Pentium 4, this limit is approximately 43 surface displacement vectors and on the 3.0 GHz Xeon, this limit is approximately 49 surface displacement vectors. Furthermore, these results have been obtained through an implementation that is not optimal and can therefore be improved upon.

4 Conclusion

In this article, we have presented a novel and general method of contact modelling. The method allows multiple arbitrarily shaped virtual tools to make contact with a virtual organ, and is independent of the size of the virtual organ

mesh. After identifying the collisions between the tools and the organ, which can be located anywhere on the surface of the organ, this field of arbitrary displacements is redistributed, using the contact modelling algorithm, into an equivalent field of displacements located only at the nodes of the mesh.

The algorithm has been tested on machines with different processors and proven to be efficient enough to meet haptic refresh rate requirements.

References

[1] C. Basdogan, S. De, J. Kim, M. Muniyandi, H. Kim, and M. Srinivasan. Haptics in Minimally Invasive Surgical Simulation and Training. *IEEE Computer Graphics and Applications,* 24(2):56–64, 2004.

[2] J. Berkley, G. Turkiyyah, D. Berg, M. Ganter, and S. Weghorst. Real-Time Finite Element Modeling for Surgery Simulation: An Application to Virtual Suturing. *IEEE Transactions on Visualization and Computer Graphics,* 10(3):314–325, 2004.

[3] S. Cotin, H. Delingette, and N. Ayache. Real-Time Elastic Deformations of Soft Tissues for Surgery Simulation. *IEEE Transactions on Visualization and Computer Graphics,* 5(1):62–73, 1999.

[4] K. Hansen, L. Brix, C. Pedersen, J. Haase, and O. Larsen. Modelling of interaction between a spatula and a human brain. *Medical Image Analysis,* 8:23–33, 2004.

[5] G. Picinbono, J. Lombardo, H. Delingette, and N. Ayache. Improving realism of a surgery simulator: linear anisotropic elasticity, complex interactions and force extrapolation. *Journal of Visualization and Computer Animation,* 13:147–167, 2002.

[6] D. Popescu and M. Compton. A Method for Efficient and Accurate Interaction with Elastic Objects in Haptic Virtual Environments. In *Proceedings of Graphite 2003,* pages 245–249, 2003.

[7] L. Raghupathi, V. Cantin, F. Faure, and M. Cani. Real-time Simulation of Self-collisions for Virtual Intestinal Surgery. In *Proceedings of the International Symposium on Surgery Simulation and Soft Tissue Modeling,* number 2673 in Lecture Notes in Computer Science, pages 15–26, 2003.

Real-Time Photo-Realistic Rendering for Surgical Simulations with Graphics Hardware

Mohamed A. ElHelw, Benny P. Lo, Ara Darzi, Guang-Zhong Yang

Royal Society/Wolfson Medical Image Computing Laboratory,
Imperial College London, London, United Kingdom
{mohammed.elhelw,benlo,a.darzi,g.z.yang}@imperial.ac.uk

Abstract: Computer-based surgical simulations are being increasingly used for training and skills assessment. They provide an efficient and cost effective alternative to traditional training methods. To allow for both basic and advanced skills assessment, the required perceptual fidelity is essential to capturing the natural behavior of the operator. The level of realism in terms of object and scene appearance determines the faithfulness and hence the degree of immersion experienced by the trainee in the virtual world. This paper presents a novel photo-realistic rendering approach based on real-time per-pixel effects by using the graphics hardware. Improved realism is achieved by a combined use of specular reflectance and refractance maps to model the effect of surface details and mucous layer on the overall visual appearance of the tissue. The key steps involved in the proposed technique are described, and quantitative performance assessment results demonstrate the practical advantages of the proposed technique.

1 Introduction

In minimal invasive surgery (MIS), virtual and augmented reality based systems are rapidly becoming an integral part of surgical training. Current high-fidelity simulators offer the opportunity for safe, repeated practice and objective measurement of performance. They provide an economical and time saving solution for acquiring, as well as assessing basic surgical skills [1]. In particular, surgical simulators are found to be valuable for training MIS procedures where the complexity of instrument controls, restricted vision and mobility, difficult hand-eye co-ordination and the lack of tactile perception require a high degree of operator dexterity [2]. Although these simulators can accelerate the development of hand-eye skills, there are serious shortcomings with the current technology, particularly in the photo-realism they provide. Hitherto, a significant amount of research has been carried out in photo-realistic rendering of simulated surgical scenes [3] and it remains one of the major technical challenges due to the complexity and diversity of internal tissue structures and surfaces properties [4].

With the recent advances in computer graphics architecture, it is possible to provide high fidelity rendering at interactive rates. Highly programmable graphics processor units (GPUs), including floating-point vertex and fragment processors, can offload complex vertex and pixel operations from the central processing unit to the

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 346-352, 2004.

GPU, allowing greater control over the graphics pipeline for real-time per-pixel shading and other procedural effects [5]. Moreover, shading calculations can be performed at the pixel level [6] as opposed to the vertex level in case of fixed functionality pipeline, hence reducing the aliasing of specular highlights and improving visual realism.

The purpose of this paper is to present a novel rendering technique based on the programmable graphics pipeline for laparoscopic simulation. Specular reflections are modulated by using a set of reflectance maps, *i.e.* maps encoding the surface normal distribution, which define the surface light interaction properties. Results are further enhanced with the improved visual appearance of the semi-transparent mucous layer on the top of tissue surface. We describe the key steps involved in the proposed technique and demonstrate its advantages over conventional approaches.

2 Method

Specular highlights constitute a vital clue in MIS procedures where 3-dimensional perception is diminished due to the use of 2-dimensional screens. Surgeons usually rely on specular highlights as a reference for depth, orientation and deformation. Consequently, it is important for surgical simulators to reproduce these highlights as realistic as possible. In the existing literature, a number of approaches for simulating specular highlights have been introduced. Standard graphics APIs such as OpenGL [7] and DirectX simulate this effect by using the Phong lighting model [8] computed at the vertices. Other techniques use environment mapping to map an image of the specular source onto the surface. The results obtained with these methods, however, generally lack visual realism due to the fact that tissue surfaces are not perfectly smooth. A more physically accurate model should consider a rough surface augmented with microstructure details such as that proposed by Torrance and Sparrow [9]. A major drawback of this approach is that physically-based reflection models can be computationally prohibitive.

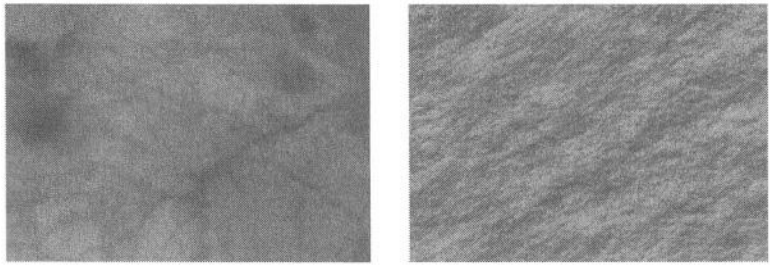


Figure 1. A sample colour texture image (left) and its associated reflectance map (right)

For real-time applications, a reflectance map can be used to describe a perturbed normal value for each image pixel (texel). This map can be either derived empirically or generated by using conventional noise functions, *e.g* Perlin noise [10]. Since the type of noise affects the shape of specular regions, different functions can be used for varying tissue types. Figure 1 illustrates a sample colour texture image and its

associated reflectance map. In this case, the reflectance map is obtained by using a noise image where every texel is considered as a height field, *i.e.* each texel encodes a single height value at that texel. The normal of the surface at each texel is then found by computing the cross product of the pair of vectors formed by that texel and its neighbors. The calculation is repeated for the other neighboring texels and the average of the normals is stored. A larger texel area can be considered in cases when smoother normals are to be computed.

During runtime, texture mapping is used for each triangle in the geometric model to extract the per-pixel reflectance map normals used for calculating the specular highlights. However, the normals in the reflectance map are defined in their own coordinate system, therefore they have to be transformed into a coordinate system that is local to the triangle being processed. Such coordinate system, known as the object local surface or texture-space coordinate system, can be defined by using three vectors which constitute its basis: the surface tangent (T), the bi-tangent (B), and the normal (N) as shown in Figure (2).

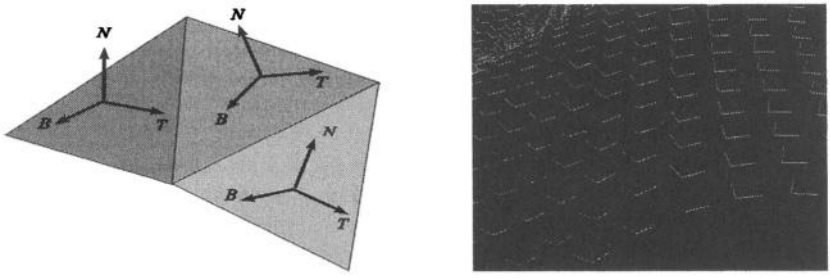


Figure 2. (Left) An example of per-triangle TBN-based coordinate systems. (Right) Per-vertex TBN bases (GBR respectively) used for the tissue model rendered in the results section.

Based on this definition, the first two vectors can be computed from the partial derivatives of the object-space coordinates of the triangle in terms of its texture coordinates [11],

$$T = \left(\frac{\partial x}{\partial u} , \frac{\partial y}{\partial u} , \frac{\partial z}{\partial u} \right) = \left(-\frac{B_o}{A_0} , -\frac{B_1}{A_1} , -\frac{B_2}{A_2} \right)$$

$$B = \left(\frac{\partial x}{\partial v} , \frac{\partial y}{\partial v} , \frac{\partial z}{\partial v} \right) = \left(-\frac{C_o}{A_0} , -\frac{C_1}{A_1} , -\frac{C_2}{A_2} \right)$$

and,

$$(A_0, B_o, C_o) = [(x_1, u_1, v_1) - (x_0, u_0, v_0)] \otimes [(x_2, u_2, v_2) - (x_0, u_0, v_0)]$$

$$D_0 = (A_0 , B_0 , C_0) \bullet (x_0, u_0, v_0)$$

where $\bullet$ denotes the dot product, (x_0, y_0, z_0), (x_1, y_1, z_1), (x_2, y_2, z_2) and (u_0, v_0), (u_1, v_1), (u_2, v_2) represent the triangular object- and texture-space coordinates respectively.

Subsequently, (N) can be computed from the cross product of (T) and (B) or the normal supplied by the original model can be used alternatively.

By computing the basis vectors in texture-space, the GPU can be used to efficiently transform the object-space vectors required for specular calculations into the texture-space by using a rotation matrix (R)

$$(R) = \begin{pmatrix} T_x & T_y & T_z \\ B_x & B_y & B_z \\ N_x & N_y & N_z \end{pmatrix}$$

Since (R) is defined for each triangle in the geometric model, a per-vertex rotation matrix is needed to ensure consistent highlights across the triangular mesh. This is obtained by averaging the rotation matrices of the triangles sharing the vertex. In practice, the rotation matrix is calculated for each vertex in pre-processing with its value at the pixel level being interpolated during the rasterisation step of the graphics pipeline, as schematically illustrated in Figure 3.

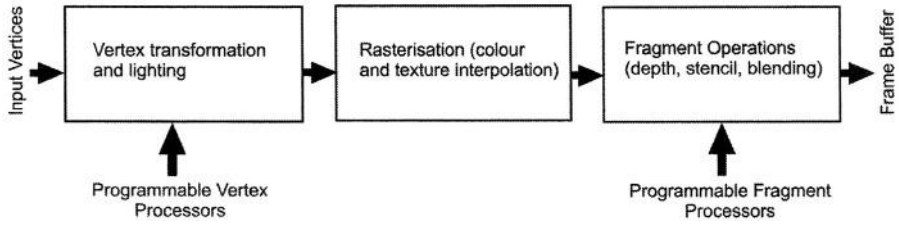


Figure 3. A simplified block diagram of the programmable graphics pipeline.

To further enhance the visual realism, the effect of surface mucous, which is a wet refractive transparent or semi-transparent layer found on top of the tissue, is combined with the above model. In laparoscopic views, the mucous layer significantly influences the surface appearance by reflecting and refracting incoming light rays. Replicating the mucous effects is a challenging problem and several factors have to be considered including the thickness of the layer, its light interaction properties, and the density and distribution of solid particles within the layer. In this study, mucous is simulated by using a set of refractance maps generated by methods similar to reflectance maps. However, vectors extracted from a refractance map are used to linearly blend between original surface colour and mucous layer colour, which accounts for surface colour variations.

3 Results

The proposed technique has been applied to endoscopic surgical simulations. A fragment shader was implemented for NVIDIA FX graphics hardware, coded in Cg [12]. Figure (4) depicts the results obtained by using the described method compared

to the conventional OpenGL multi-texturing approach. It is evident that the method effectively avoids the problem of plastic-like surface and provides realistic specular highlights. Furthermore, by varying the colour of the mucous layer and using different noise types, tissue appearance can be modified. Figure (5) demonstrates the effect of different noise functions on the visual appearance of the rendered surface.

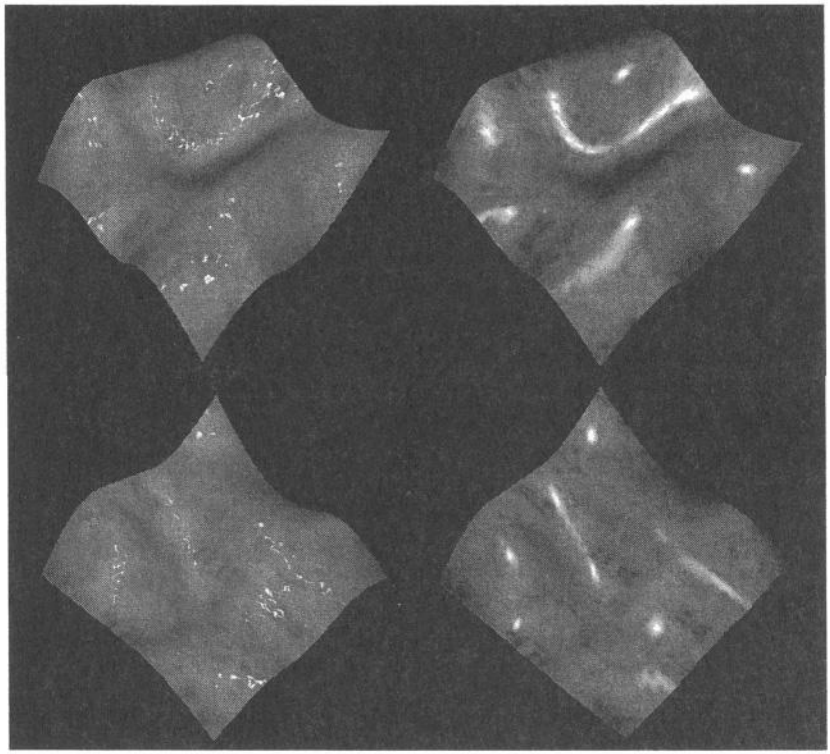


Figure 4. Different views the surface rendered by using the proposed method (left) versus OpenGL multi-texturing approach (right). Notice the plastic-like surface and the hexagonal shape of the specular highlights with the multi-texturing method.

To assess the overall computational burden of the proposed algorithm, a detailed performance analysis was carried out. The effect of using different viewport resolutions and polygon counts on performance is demonstrated in Figure 6. It is shown that the viewport resolution is inversely proportional to the achieved frame rate, which is due to the fact that the fragment program is executed for each rendered pixel. This problem can be alleviated by future graphics hardware with more fragment pipelines. Increasing the scene polygonal count, on the other hand, has a gradual impact on performance unless extensive vertex programs are used. In fact, the performance in the case of programmable graphics hardware is dependant on several factors such as the length and complexity of vertex and fragment programs and the amount of data transferred between the CPU and GPU each cycle.

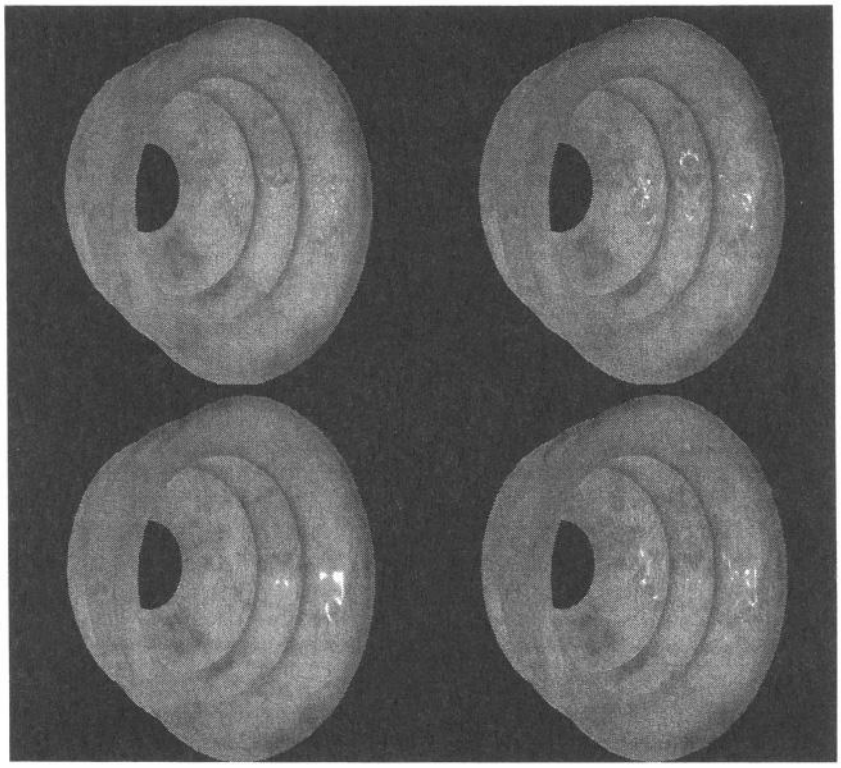


Figure 5. The effect of different noise functions on the overall visual appearance of the rendering results. Shown above are four types of noise with decreasing frequency (clockwise from top-left)

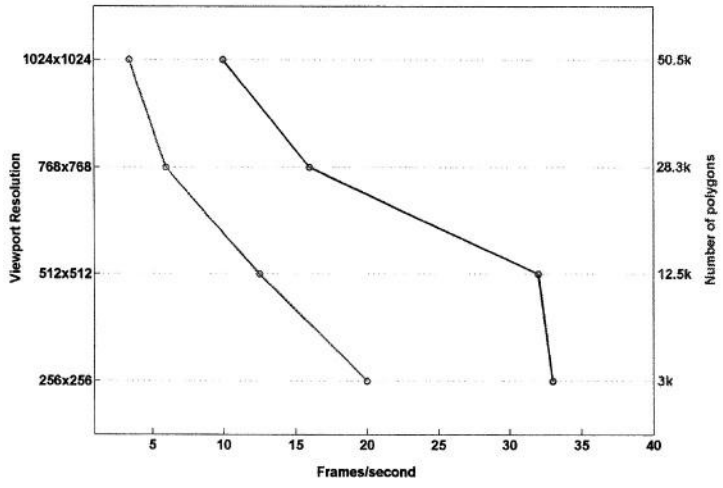


Figure 6. Performance assessment of real-time per-pixel shading with graphics hardware for different viewport resolutions and polygon counts.

4 Discussions and Conclusion

In this paper a novel photo-realistic rendering method suitable for surgical simulation is described. It is based on using combined reflectance and refractance maps to model the effect of surface details and mucous layer on the overall visual appearance of the rendering results. The programmable graphics hardware is used to allow for per-pixel control and to carry out most of the required computations. In addition to the high fidelity rendering results achieved, the computational performance achieved makes it suited for interactive MIS simulation. With the use of general-purpose capabilities of the GPU, it is also possible to migrate simulation tasks such as collision detection and deformation computation to the GPU [13], allowing the entire simulation system to be seamlessly integrated.

References

1. Shah, J. and Darzi, A.: Simulation and Skills Assessment. International Workshop on Medical Imaging and Augmented Reality (MIAR '01), Hong Kong (2001) 5-9
2. Bro-Nielsen, M.: Simulation Techniques for Minimally Invasive Surgery. Journal of Minimally Invasive Therapy & Allied Technologies (1997) 106-110
3. Neyret, F., Heiss, R. and Senegas F.: Realistic Rendering of an Organ Surface in Real-Time for Laparoscopic Surgery Simulation. The Visual Computer. Vol. 18 No. 3 (2002) 135-149
4. Stylopoulos, N., e.t al: CELTS: A Clinically-Based Computer Enhanced Laparoscopic Training System. Medicine Meets Virtual Reality (2003)
5. Peercy, M.S., Olano, M., Airey, J., and Ungar, J.: Interactive Multi-Pass Programmable Shading. Proceedings of SIGGRAPH (2000) 425-432
6. Eveirtt, C.: Mathematics of Per-Pixel Lighting. http://developer.nvidia.com /object/ mathematicsofperpixellighting.html
7. Neider J., Davis, T. and Woo, M.: OpenGL Programming Guide. 2nd edn. Addison Wesley (1997)
8. Phong, B., T.: Illumination for Computer Generated Pictures. Communications of the ACM, Vol. 18 No. 6 (1975) 311-317
9. Torrance, K.E. and Sparrow, E.M.: Theory for Off-Specular Refelction from Roughned Surfaces. Optical Society of America, Vol. 57 No. 9 (1976) 1105-14
10. Ebert, D.S. (ed.): Texturing and Modelling: A Procedural Approach. AP Professional (1994)
11. Fernando, R. and Kilgard, M.J.: The Cg Tutorial. Addison Wesley (2003)
12. Mark, W.R., Glanville, R.S., Akeley K. and Kilgard, M.J.: Cg: A System for Programming Graphics Hardware in a C-Like Language. Proceedings of SIGGRAPH (2003) 896-907
13. Fernando, R. (ed.): GPU Gems. Addison Wesley (2004)

Computer-Assisted Evaluation of Double-Bundle ACL Reconstruction

Stefano Zaffagnini, Sandra Martelli, Marco Bontempi, Simone Bignozzi

Laboratorio di Biomeccanica, Istituti Ortopedici Rizzoli, 40136 Bologna, Italy
http://www.ior.it/biomec/
{s.zaffagnini, s.martelli, m.bontempi, s.bignozzi}@biomec.ior.it

Abstract. This paper reports a new study for comparing two double-bundle reconstructions of anterior cruciate ligament (ACL), used by one of the authors in surgery. It describes the experimental setup, protocol and results in a sample case. The experiment was performed in five steps, using FlashPoint optical system: first the intact knee kinematics was analysed; then ACL was resected and the knee examined; ACL was reconstructed using the gracilis semitendinous tendon and the new kinematics recorded; ACL was reconstructed again with the same tendon and a different orientation of the femoral tunnel and kinematics was recorded; finally the knee was dissected to digitize bone surfaces, ligaments insertions and tunnel position. The off-line elaboration of data allowed to evaluate anatomical and functional results. The comparison of reconstructed knee laxities and ranges of motion with normal and ACL-deficient knee suggested that the most vertical tunnel performed better than the horizontal one.

Keywords: ACL reconstruction, knee, kinematic analysis, computer analysis

1 Introduction

Anterior Cruciate Ligament (ACL) reconstruction is a frequent intervention. Recently to improve the clinical results of this procedure same authors have presented a more anatomical reconstruction trying to reproduce the two bundles of the normal ACL. This reconstruction should have a better performance from the kinematic point of view. However very few studies have analysed the kinematic performance of such procedure, especially the bundles features affecting the final outcome. Therefore a more and more careful analysis of the double-bundle surgical technique is needed to be able to predict and optimize the functional results of the procedure.

In this study we present a new method to analyse the double-bundle ACL reconstruction, based on the use of a navigation system for data acquisition (i.e. for tracking motion and digitizing anatomical surfaces) and successive computer elaboration of the kinematic and anatomical results. The accuracy of the system allows a reliable analysis also of secondary kinematic constraints and, on a cadaveric specimen, also a reliable study of the relationship with anatomical features [1, 2].

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 353-360, 2004.

This study describes the acquisition and elaboration protocol of the methodology and reports a case study to investigate the effect of tunnel orientation in double-bundle ACL reconstruction on the kinematics of the reconstructed knee.

2 Materials and Methods

2.1 Materials

For the analysis of the double-bundle ACL reconstruction in one cadaver knee we used an optical navigation system (FlashPoint, Image Guided, Boulder, Colorado) to record relative motion of the tibia and the femur and to digitize anatomical data. This device has been used by several authors in surgical and laboratory tests of computer-assisted interventions and all authors have confirmed its sub-millimetric accuracy [3].

The femur was fixed to the experimental desktop with a clamping device, while the tibia was left free to move like in the standard operating room set-up. The foot was intact and let the surgeon check the internal-external alignment of the limb like in the operative room. The femur was fixed horizontally with tibia, at 90° of flexion, perpendicular to the floor in order to minimize external forces due to the weight of the leg and have a physiological passive range of motion. This setup also let the surgeon have an easy access to the internal part of the knee during the ACL reconstruction.

Two rigid bodies with infra-red emitters were fixed respectively to femur and tibia in order to record their relative position during passive motion. The setup was optimized to minimize possible interactions with the surgical actions, therefore the femoral rigid body was fixed in the proximal part of the femur and the tibial one distally in the medial part of the tibia.

2.2 Acquisition Protocol

The acquisition protocol of our study consisted in six main steps: four steps aimed at acquiring information about the kinematic behaviour of the intact, ACL-deficient and reconstructed knees with two different techniques; one step was the execution of standard surgical technique for double bundle ACL reconstruction, and a last step concerned the acquisition of anatomical data, in order to have a complete analysis of the cadaveric joint. In details the acquisition protocol was the following [4]:

- Kinematic acquisitions on the intact knee,
- ACL minimally invasive ACL recision,
- Kinematic acquisitions on the ACL-deficient knee,
- Gracilis and semitendinous tendons were harvested and sutured together with the tibial insertion left intact. A tibial tunnel was performed in a way to reach the natural tibial insertion area of ACL, in particular its postero-medial part. The tunnel started in the medial part of cresta tibialis and had an orientation of 7° with respect to the anatomical axis in the frontal plane and 37° in the sagittal plane [5].
- Preparation of femoral tunnel at 2.30 o'clock [6], corresponding to 28° with respect to the tibial plateau in flexion, and ACL reconstruction (called "horizontal tunnel" in the following). Tendons passed through the tibial tunnel, over the top, in the

horizontal tunnel (HT) and in the tibial tunnel again, and clamped with a barred staple to have a rigid ligament fixation;
- Kinematic acquisitions on the knee reconstructed with double-bundle technique and horizontal tunnel orientation;
- Preparation of femoral tunnel at 1 o'clock [6], corresponding to 37° with respect to the tibial plateau in flexion, and ACL replacement (called "vertical tunnel" in the following). Tendons passed through the tibial tunnel, over the top, in the vertical tunnel (VT) and in the tibial tunnel again, and clamped with a barred staple to maintain ligament fixation
- Kinematic acquisitions on the knee reconstructed with double-bundle technique and vertical tunnel orientation;
- Acquisition of standard joint coordinate reference system;
- Knee dissection and ligaments' exposure;
- Digitization of ligaments' insertions, tunnel entrance and exit holes on the bones, distal femur shape and proximal tibia surface.

Kinematic acquisitions included the passive range of motion from full extension to full flexion, the internal/external rotation at 90° of flexion and at maximum force, and the drawer test at maximum force. All of them were recorded twice by the same surgeon and all the 6 degrees of freedom of the knee joint were recorded during tests in different conditions.

2.3 Computer Analysis

The computer analysis of the knee joint and the kinematic data was performed with a custom software that allowed the reconstruction of relative motion of the joint and the computation of laxities, as well as instantaneous rotations and translations [1].

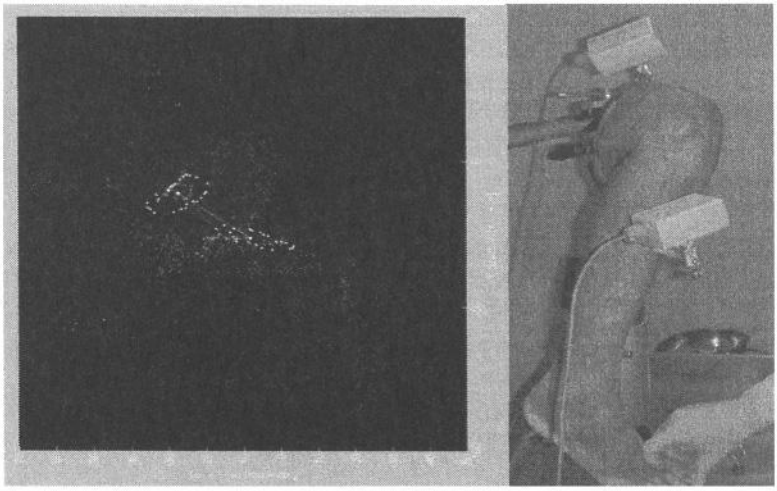


Fig. 1. Display of the anatomical reconstruction aligned to the X plane (left) and experimental setup with femur fixed on desktop and FlashPoint's frames attached to femur and tibia (right).

Transepicondylar line and mechanical axis on femur and medial lateral direction and anatomical axis on tibia were acquired and used as local reference coordinate systems for the two bones. The display joint reference system was adjusted so that the femoral condyles profiles coincide in the sagittal view and the line tangent to femoral condyles in flexion was horizontal (Fig. 1).

For kinematic tests the joint reference system was defined as follows: the x axis was fixed to the femur, and defined as the transepicondylar line; the z axis was fixed to the tibia, and defined as the long tibial axis normalized with respect to the x axis; the y axis was the result of the instantaneous cross product of the z and x axes. In our experiment the femur and the x axis were fixed while the z and the y axes were mobile.

The decomposition into flexion-extension (FE) rotation, internal-external (IE) rotation and varus-valgus (VV) rotation was computed using cardan angles in the sequence X-Z-Y. Similarly translations were computed along these floating frames at each flexion angle. In particular for the elaboration of kinematic data we computed:

- The amount of FE, IE and VV rotation during the passive range of motion (PROM) in the four different knee states;
- The FE, IE and VV rotation at the position recorded for stress tests in the four different knee states (called knee "attitude");
- The antero-posterior (AP) laxity of the joint, as the maximum 3D displacement occurring during drawer test;
- The internal-external (IE) laxity of the joint, as the helical angle between the most medial and most lateral position attained during the IE rotation test at 90°;
- The elongation and orientation of the linear fibre joining the centres of the insertion areas of the original ACL during kinematic tests;
- The elongation of the two bundles of the reconstructed ACL during PROM in reconstructions with horizontal and vertical tunnel;
- The orientation of the two bundles of the reconstructed ACL during PROM with respect to tibial plateaux in reconstructions with horizontal and vertical tunnel.
- The orientation of the two bundles of the reconstructed ACL during PROM with respect to femoral notch in reconstructions with horizontal and vertical tunnel.

For all kinematic computation we computed the "error" associated to the measurement as the standard deviation of the mean of repeated motions.

3 Results

The passive range of rotations during PROM were similar in intact, ACL-deficient and reconstructed knees (Table 1), although small differences were found in the initial attitude of the joint in full extension.

The position for the drawer tests appeared slightly (i.e. not statistically significant) different in the AP (i.e. y) and medio-materal (i.e. x) directions, but similarly oriented in all knee conditions, as shown in Table 2.

AP and IE laxities varied according to the ACL state, as reported in Table 3 as the mean among all repeated tests in the same conditions.

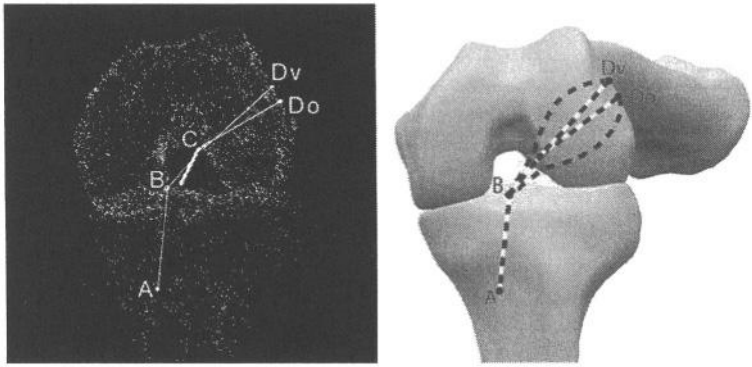


Fig. 2. Bundle reconstruction at 90° of flexion. Software display (left) and schematic representation (right). A: external insertion of tibial tunnel, B: internal insertion of tibial tunnel, C: internal insertion of vertical and horizontal femoral tunnel, Do: external insertion of femoral horizontal tunnel, Dv: external insertion of femoral vertical tunnel.

Table 1. Rotations occurring during PROM in the four knee conditions

	Intact	no-ACL	Hor. Tunnel	Vert. Tunnel
Flexion	85°	90°	93°	92°
IE	-20°	-22°	-18.5°	-18°
VV	-5°	-6°	-6°	-6°

Table 2. Attitudes of the knee at IE test (IE) and drawer test (AP)

	Intact			NO-ACL			H Tunnel			V Tunnel		
	Flex	IE	VV	Flex	IE	VV	Flex	IE	VV	Flex	IE	VV
AP	77°	-17°	-3.5°	82°	-17°	-3.5°	80°	-18°	-4°	82°	-18°	-4°
IE	82°	-15°	-5.5°	80°	-15°	-5.5°	81°	-15°	-5°	78°	-15°	-4.5°

Table 3. Knee laxities

	AP laxity (mm)	IE laxity (deg)
Normal	32.1 ± 1.7	25.8 ± 0.2
No-ACL	46.6 ± 2.0	29.4 ± 1.2
HT graft	41.6 ± 1.6	20.4 ± 0.5
VT graft	31.4 ± 1.2	28.1 ± 0.4

Legenda: Normal = Intact knee; **No-ACL** = ACL-deficient knee; **HT graft** = knee reconstructed using double-bundle technique with horizontal tunnel orientation; **HT graft** = knee reconstructed using double-bundle technique with vertical tunnel orientation.

Notice that the tendons length in our specimen was enough to reach the final clamping site for both femoral tunnel orientation. The final length in the two reconstructions can be computed at extension from the sum of the following length:

1. Twice the tibial tunnel length, which resulted in 48.9 mm;
2. The length of the bundle of the reconstructed knee going from the tibial tunnel to the femoral tunnel in both cases, that we will call "antero-medial" bundle of the reconstructed ACL, equals to 27 mm for HT and 28 mm for VT in flexion;
3. The femoral tunnel length, equal to 35.8 mm for HT and 42.1 mm for VT;
4. The length of the tendon wrapping around the posterior lateral condyle under the capsula, equal to 32.9 mm for HT and 39.1 mm for VT;
5. The length of the bundle from tibial tunnel to over the top, that we will call "postero-medial" bundle of the reconstructed ACL, equals to 35 mm for HT and 36.2 mm for VT in flexion.

As the position of both bundles of the reconstructed ACL with HT and VT was very similar (and inserted laterally in the tibial and distally in the femoral tunnel), the difference of the graft length in the two techniques was due to item 3. and 4. and was a fixed offset of 12.5 mm at extension. The total length of the graft during PROM and stress tests varied, mainly due to the elongation of the antero-medial (item 2.) and postero-medial (item 5.) bundle of the graft.

Their elongations are reported in Fig.3. Both the anterior bundles of the reconstructed ACL resulted isometric during PROM, like the normal ACL central fibre, while both the posterior bundle decreased its length in flexion by almost 20%.

Also the orientation of the reconstructed ACL bundles were similar both in the horizontal and vertical case, and differed the central fiber of the natural ACL, especially in extension. The orientation of the anterior bundles of the reconstructed ligament with respect to the tibial plateau was similar to the normal one and decreased in flexion. The orientation of the posterior bundle of the reconstructed ACL with respect to the tibial plateau varied much more during PROM.

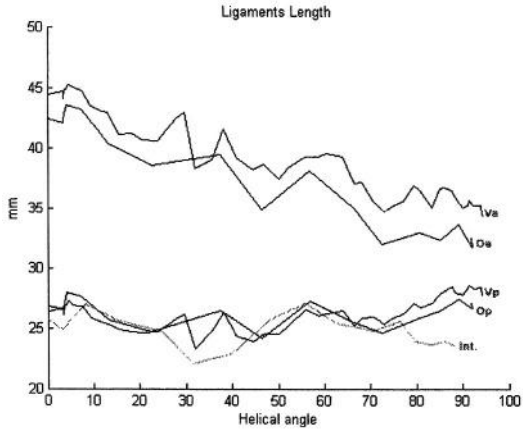


Fig. 3. Length variation of intact and reconstructed ACL during PROM. Int: **ACL** (line joining centres of insertions) **O**: Horizontal tunnel bundles (a= anterior p=posterior) **V**: Vertical tunnel bundles (**a**= anterior **p**=posterior)

The orientation of both anterior and posterior bundles of the reconstructed ACL with respect to the femoral notch were quite similar, and increased less than normal ACL during PROM.

4 Discussion and Conclusions

The experimental setup we have proposed was able to provide a significant comparison among normal knee, ACL-deficient knee and reconstructed knee with two double-bundle techniques. The most original part of this study is the new protocol for the comparison of the 3D kinematics of the reconstructed knees and the normal and ACL-deficient one in a single specimen (which avoids the problem of individual variability). The kinematic comparison is quite original and includes information rarely reported in previous studies, such as graft orientation and full length.

We can notice that the analyses of the passive range of motion or stress tests (Table 1, 2) did not discriminate between the reconstruction techniques with HT and VT, and was not even able to distinguish normal and ACL-deficient knees. This is consistent with previous studies in literature [7], and is probably due to the fact that ACL does not act as an active constraint during PROM and passive motions.

An encouraging observation on the two reconstructions is that either HT or VT reconstructions were able to restore the ACL elongation and general orientation of the ligament [2,8]. This is probably due to the fact that the tibial tunnel appeared within the area of the original ACL tibial insertion, near ACL posterior bundle, and therefore the common anterior-bundle of two graft reconstructions contributed significantly to restore the normal knee passive kinematics.

However a difference can be noticed between HT and VT when comparing the reconstructed knee laxities. In fact only the vertical tunnel was able to restore the natural AP and IE stability of the knee, while the horizontal one appeared unable to fully control AP laxity and constrained IE rotation more than the normal ACL.

This result suggest that the femoral tunnel orientation may have a significant effect on the final knee behaviour during stress test, as already described by Woo [6,9], even if no significant differences could be noticed during passive motion. This difference may be due to the different global length of the used tendon or, although small, different positions of the antero-medial or postero-medial bundle of ACL reconstructions. In fact all the other features of the reconstructions were the same in the two examined technique (i.e. the graft, the tibial tunnel, the position of the anterior and posterior bundle of the reconstructed ACL).

We remark that this result appears surprisingly different from the classic single bundle technique, where the control of AP laxities increases when the orientation of the femoral part becomes more horizontal [6]. This result may be due to the physical behaviour of the tendon wrapping around the femoral condyle with different length and therefore forces in the two double-bundle techniques and absent in the single-bundle one.

We plan to perform a more extensive experimental study of this issue in the near future.

References

1. Martelli S: New Method for simultaneous anatomical and functional studies of articular joints and its application to the human knee, Comp Methods & Programs in Biomed, 70: 223-240, 2003 .
2. Zaffagnini S, Martelli S, Acquaroli F: Computer investigation of ACL orientation during passive range of motion. Computers in Biology and Med, 34:2, 2004 (to appear)
3. Martelli S, Bignozzi S, Bontempi M, Zaffagnini S, Garcia L.: Comparison of an optical and a mechanical navigation system. R.E. Ellis and T.M. Peters, MICCAI 2003, LNCS 2879, 303-310, 2003
4. Martelli S, Zaffagnini S, Falcioni B, Motta M.: Determination of an optimal kinematic protocol for computer-assisted evaluation of anterior cruciate ligament deficiency. Ann Biomed Eng. 2001 Dec;29(12): 1112-21.
5. Marcacci M, Molgora AP, Zaffagnini S, Vascellari A, Iacono F, Presti ML: Anatomic double-bundle anterior cruciate ligament reconstruction with hamstrings. Arthroscopy. 2003 May-Jun;19(5):540-6
6. Loh JC, Fukuda Y, Tsuda E, Streadman RJ, Fu FH, Woo SL: Knee stability and graft function following anterior cruciate ligament reconstruction: Comparison between 11 o'clock and 10 o'clock femoral tunnel placement, Arthroscopy. 2003 Mar;19(3):297-304.
7. Jalliard R, Lavallee S, Dessenne, V: Computer assisted reconstruction of the anterior cruciate ligament. Clinical Orthopedic. 1998 Sep;(354):57-64.
8. Sapega, AA, Moyer, RA, Schneck C, Komalahiranya, N: Testing for Isometry during reconstruction of the anterior cruciate ligament, Journal of Bone and Joint Surgery 72-A[2]:259-267, 1990
9. Mushal V, Burkart A, Debski RE, Van Scyoc A, Fu FH, Woo SL: Anterior cruciate ligament tunnel placement: Comparison of insertion site anatomy with the guidelines of a computer-assisted surgical system, Arthroscopy. 2003 Feb; 19(2): 152-60.

Integral Videography Overlay Navigation System Using Mutual Information-Based Registration

Hongen Liao, Takashi Inomata, Nobuhiko Hata, and Takeyoshi Dohi

Graduate School of Information Science and Technology, the University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
liao@atre.t.u-tokyo.ac.jp

Abstract. We propose an automatic markerless registration method for Integral Videography (IV) overlay navigation system that directly matches a video image of the patient to a 2-D rendering of a surface model extracted from the 3-D surface model image. IV overlay system is integrated into a surgical navigation system to superimpose a real three-dimensional (3-D) autostereoscopic image onto the patient via a half-silvered mirror. This registration technique doesn't require the setting of markers, which is invasive and extends the duration of the procedure. Accuracy measurements were performed using a triangular-prism-shaped model. The mean Target Registration Error (TRE) was 1.6 ± 0.6 mm. We also applied our method to a human face registration and obtained successful results, although some noises were added to the video image. The presented method is more suitable for clinical applications of IV image overlay than previous methods since the presented method is sufficiently accurate, robust and more convenient for surgeon.
Keyword: Image overlay, navigation, 3-D image, mutual information.

1. Introduction

Image overlay system is surgical navigation that superimposes medical images over surgical fields on the patient. Surgeons can see beneath the surgical scene and view hidden structures as if they could be seen through the body. By localizing the targeted lesion and the critical lesion that should be avoided, the image overlay navigation helps to achieve effective and safe surgery while minimizing the invasiveness of the surgery. This overlay function eliminates hand-eye coordination issue often raised in navigation surgery, where image display is placed off the operative field and forces the physician to look away from the surgical field.

The use of a head-mounted display (HMD) is one of the most important innovations because it augments the surgeon's view of the surgical field with computer-generated images [1]–[3]. Birkfellner *et al.* presented a basic design of a modified HMD to increase the clinical acceptance of augmented reality [2]. Fuchs *et al.* reported a 3-D visualization system with an HMD for use in laparoscopic surgical procedures [3]. The HMD is tracked to compute an appropriate perspective. These systems still have a problem of lag for motion parallax and cannot provide a natural

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 361-368, 2004.

view for multiple observers. These techniques have the potential to enhance the surgeon's ability to successfully perform complex procedures.

Another type of image overlay system uses a half-silvered mirror display device, in which medical images are superimposed over the patient in the real world. Blackwell *et al.* introduced an image overlay system that uses a binocular stereoscopic vision display and describes an image overlay prototype with initial experimental results [4]. By using image overlay with 3-D reconstructed images, surgeon can visualize the data "in-vivo" when it is aligned exactly to the patient's anatomy.

We have developed an autostereoscopic image overlay technique that can be integrated into a surgical navigation by superimposing an actual integral videography (IV) image onto the patient by use of a half-silvered mirror [5]. IV record and reproduce 3-D images using a micro convex lens array and a flat display. IV uses both horizontally and vertically varying directional information and thus produces a full parallax image, making it a promising method for creating 3-D autostereoscopic displays. With additional improvements in the display, this system can increase the surgical accuracy and reduce invasiveness. In our system, image-to-physical registration is essential and markers are used for this purpose. When using markers, there is a need to attach markers to the patient, which is invasive. In addition, markers' positions must be measured using an optical tracker, which extends the duration of the procedure. Nicolau *et al.* developed an augmented reality system using markers without optical trackers [6].

The objective of this study is to perform an automatic markerless registration of the graphical model and image overlay device using maximization of mutual information. Specifically, our method is an extension of mutual-information based video-to-model registration [7]. Without using markers, the newly proposed method is clinically significant since it is non-invasive for patients and less cumbersome for medical staff. The engineering contribution of this study is the use of mutual information for patient to 3-D autostereoscopic image registration in image overly system, which in our knowledge has not been reported elsewhere.

2. Materials and Methods

To perform an automatic markerless registration for IV image overlay systems, we directly matched a video image and a 2-D rendering of a surface model extracted from 3-D image. The goal of image-to-physical registration in image overlay systems is to determine the transformation matrix of display-to-patient and register the IV image to patient's body.

2.1 IV Image Overlay Systems

The IV display we developed consists of a high-resolution LCD with a micro convex lens array, a half-silvered mirror, and a supporting stage. IV records and reproduces 3-D images using a micro convex lens array and flat display; it can display geometrically accurate 3-D autostereoscopic images and reproduce motion parallax without the need for special devices. The use of semi-transparent display devices makes it appear that the 3-D image is inside the patient's body.

Generally, the surgeon observed the operating field directly by eye during the operation. The fundamental elements of this system include 3-D image on IV display, CCD camera, and patient. The relationships of these elements are shown in Fig. 1.

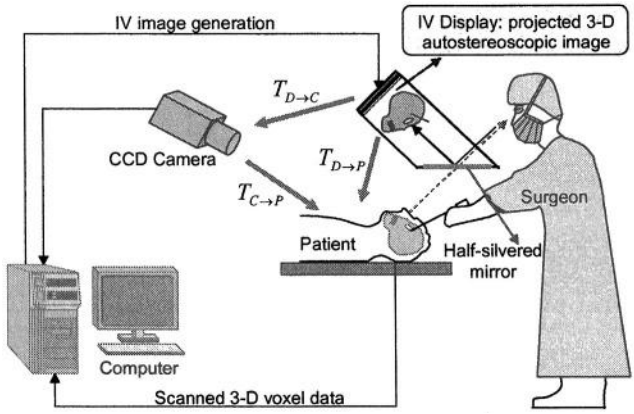


Fig. 1. System configuration: instrumentation for automatic surgical navigation based on IV.

2.2 Coordinate Transformation

The coordinate transformations required to produce an overlay image that is registered to the patient are $T_{D\to P}$, $T_{D\to C}$ and $T_{C\to P}$, where *C, D* and *P* stand for CCD camera, display and patient, respectively.

Figure 1 shows each coordinate transformation. To register an overlay image to a patient, a transformation matrix of display-to-patient, $T_{D\to P}$, is required. The matrix is calculated with following equation:

$$T_{D\to P} = T_{D\to C} T_{C\to P} \tag{1}$$

The transformation matrix of display-to-camera $T_{D\to C}$ is constant and obtained before registration. The transformation matrix of camera-to-patient $T_{C\to P}$ is determined by matching a video image of a patient to a 2-D rendering extracted from the 3-D surface model.

2.3 Optimization of Matching

To determine the transformation matrix of camera-to-patient, it is necessary to find the transformation matrix that maximizes the similarity measure, using following equation:

$$\hat{T} = \arg\max_{T} f\left(u(x), v(T(x))\right) \tag{2}$$

where x is a random variable over coordinate locations in the image, u and v are intensities of each image, f is the similarity measure, and T is the transformation

matrix of camera-to-patient. We found the optimal transformation matrix $\hat{T}_{C\to P}$ by gradient descent optimization:

$$T_{next} \leftarrow \lambda \cdot \frac{d(f)}{d(T)} + T \tag{3}$$

where λ is a positive constant called learning rate. The transformation matrix that maximizes the similarity measure is determined as an optimization proceeds. Calculation of the similarity measure, gradient and reprojection of the surface model are repeated until the number of step in the optimization reaches a defined number of iterations.

2.4 Similarity Measure Using Mutual Information

We used Mutual Information (MI) as a similarity measure. MI is generally used in information theory. It can be applied to image registration [8]. When thus applied, MI is measured as the statistical dependence, or information redundancy, between the image intensities of corresponding pixels in both images.

MI of two images denoted by *I(u(x),v(T(x)))* is given as the following equation:

$$I\big(u(x), v(T(x))\big) = H\big(u(x)\big) + H\big(v(T(x))\big) - H\big(u(x), v(T(x))\big) \tag{4}$$

where $H(\ )$ is the entropy of a random variable and can be interpreted as a measure of uncertainty. The value of MI becomes high when one image is similar to the other. MI can be used to align images of different modalities, which is suited to matching a video image to a 2-D rendering. In addition, MI has high robustness which allows us to resolve problems of the lighting source of operating rooms and noises, such as hair and eyes, which are not in 3-D surface models but are in video images.

2.5 Registration Process

The following scenario describes our registration method in image overlay systems.

1) *Construct the surface model:* A patient is scanned by a 3-D internal anatomy scanner, such as magnetic resonance imaging (MRI) or computerized tomography (CT). The surface area of the patient is segmented manually.
2) *Set up image overlay devices:* The patient is placed in the operating room and fixed on the operating table. Image overlay devices, which include a CCD camera, are set up over the patient. The relative positions of each device are fixed.
3) *Determine the initial position and parameters:* Prior to draping, the image of patient is taken by the CCD camera. Two images, the video image and the 2-D rendering image of the surface model extracted from the 3-D model constructed in 2), are aligned manually. This determines the initial position of the optimization. In addition, parameters such as learning rate and iteration number are determined.
4) *Optimize MI:* Optimization of MI is achieved as shown in Equation 3. Optimization proceeds through a defined number of iterations. The surface model is reprojected at each iteration. The transformation matrix $T_{C\to P}$ is obtained.

5) *Obtain the Overlay IVImage:* The transformation matrix $T_{D \to P}$ is calculated from Equation 1. The patient is draped and the overlay IV image is displayed. Surgeons can see the overlay image during the surgery as if it was within the patient.

3. Experiments and Results

3.1 IV Image Overlay Systems

We performed three experiments to assess the accuracy, processing time and suitability to clinical applications of our approach. The registration method was implemented on an IV navigation system. We calculated the value of MI and gradient using Insight Segmentation and Registration Toolkit (ITK). A Visualization Toolkit (VTK) was used to obtain a 2-D rendering surface model.

We used an IV navigation system for image overlay (Fig. 2). The pitch of each pixel on the screen of IV display is 0.120 mm. Each lenslet element is square with a base area of 1.001×0.876 mm, which covers 8×7 pixels of the projected image.

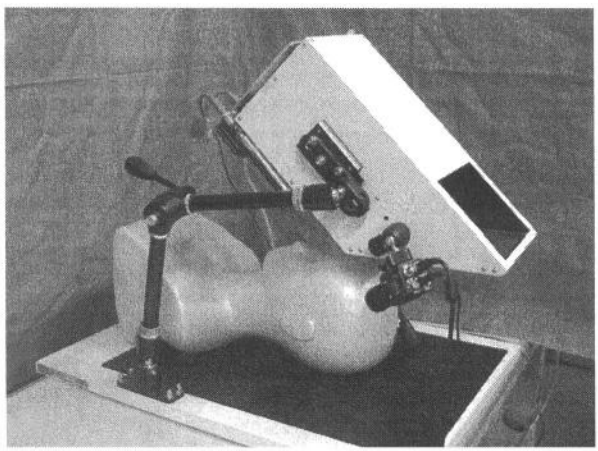


Fig. 2. IV overlay navigation system with automatic image registration, including the IV display, the half-silvered mirror, and the CCD camera that fixed with the same hardware.

3.2 Registration Accuracy Experiment and Processing Time Measurement

We used a triangular-prism-shaped phantom to measure the registration error easily. The position of four vertexes of the overlay image and the phantom are measured by use of a POLARIS optical system (Northern Digital Inc. Ontario, Canada). The vertex positions were measured three times and the mean is considered as its coordinate value. The four vertexes were on different planes in order to assess the registration error three-dimensionally. We assumed that there was a target, such as a tumor, in the centroid of the phantom, whose position was obtained from the vertexes. We also calculated the distance between two targets as the target registration error (TRE) [9].

The phantom was placed in four different positions and the TRE of each position was calculated.

We measured the processing time of the registration by use of a standard Linux PC (CPU: Pentium4 2.53 GHz, RAM: 1024 MByte). Parameters of MI were constant during the experiment. We made the iteration number high enough to obtain the convergence of MI.

The TREs of registration were measured in six positions for 10 times measurement. The mean TRE was 1.6±0.6 mm (mean ±SD). The processing time was measured to be 60 ±1 seconds (mean ±SD, n=10).

3.3 Matching on Human Face

For evaluating the feasibility of clinical applications, we matched a video image of a human face to an extracted 2-D rendering surface model. A MRI data (T1 weighted, Matrix: 128×128×128, Voxel Size: 1.9×1.9×1.5 mm) is used to construct the 3-D surface model. We registered the 2-D rendering surface model to the video images manually and implemented the optimization. The 2-D rendering surface model is matched to the video images of the real human face in spite of noises such as hair and eyes. Figure 3 shows the result of matching a video image to a human face. The video image was successfully matched to the 2-D rendering surface model extracted from the 3-D surface model by maximizing the similarity measure.

The results of MI convergence are shown in Table 1, which show maximal offset and rate of the successful convergence. In addition, final translation and rotation are also given in this table.

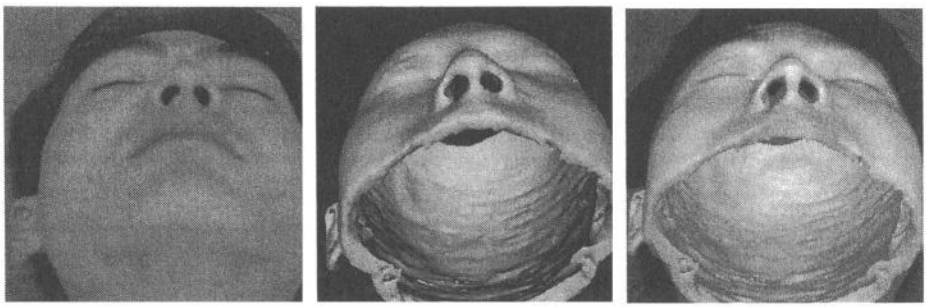


Fig. 3. Result from the matching experiment on a human face. (a) CCD image. (b) Model image, (c) Final overlay image. The overlay images consist of a CCD image and a model image. The final overlay image shows the image resulting from the optimization.

3.4 Clinical Studies of IV Image Overlay

Intra-operatively, IV autostereoscopic image can help with the navigation by providing a broader view of the operation field by use of an IV overlay technique. We

Table 1. Results of MI convergence.

Experiment	Maximal offset		Result error		Success rate
	Trans (mm)	Rot (deg)	Trans (mm)	Rot (deg)	(%)
# 1	10	0	0.9	0.02	100
# 2	20	0	1.1	0.03	90
# 3	30	0	5.7	0.04	70
# 4	0	10	1.0	0.03	100
# 5	0	20	1.2	0.05	93
# 6	0	30	3.8	0.11	73
# 7	20	20	0.9	0.03	100
# 8	30	20	4.4	0.05	93
# 9	20	20	3.2	0.14	73

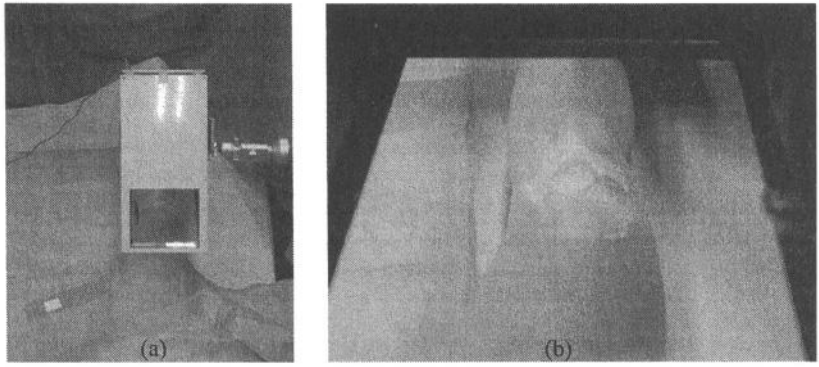


Fig. 4. Clinical application experiment of IV image overlay. (a) IV image overlay device (b) Surgical implementation integrated with IV image for knee arthroplasty.

superimposed IV image into the patient in surgical implementation and integrated IV image overlay in knee arthroplasty (Fig. 4). These combinations enabled safe, easy, and accurate navigation. In combination with robotic and surgical instrument, it even can supply guidance by pre-defining the path of a biopsy needle or by preventing the surgical instruments from moving into critical regions.

4. Discussions and Conclusion

The result of error measurements indicates that our method is sufficiently accurate to be used for clinical applications, compared with a mean target registration error of approximately 3 mm for skin marker registration in clinical practice [10].

However, it is necessary to improve the registration accuracy since the registration error would be increased in the case of real patients. The CCD camera sensitivity to rotation in its depth is considered as a factor causing the error. Registration accuracy can be improved by using multiple CCD cameras. However, the use of multiple cameras increases the processing time. Further investigation is needed on this subject.

The processing time of our registration method was about 60 seconds. It is comparatively rapid contrast to the registration using markers. Furthermore, our method is less invasive than that of using markers since it is no need to attach markers to the surface of patients. The processing time could become much shorter if we adjust the optimal parameters of the similarity measure. However, adjusting the parameters is complicated and took a long time. It is a challenging work to find optimal parameters automatically.

The experiment of matching on a human face was successful. We obtained successful convergences of MI and confirm the robustness of MI, although there were noises in the special part such as hair and eyes. Since the noises were not in surface models but in video images, the experimental results strongly suggests that our method can be applied to real patients. In future work, we will assess the registration accuracy on a clinical implementation and introduce it the field of orthopedic surgery like the operation of knee and wrist.

In conclusion, we have developed an automatic markerless registration method for IV autostereoscopic image overlay systems that directly matches a video image of a patient to a 2-D rendering surface model extracted from the 3-D surface model. In experiments, the TRE was 1.6 mm and the processing time was approximately 60 seconds. The results suggest that our method is sufficiently accurate to be used for clinical applications and the processing time is much shorter than that of using markers. We also success applied our method to human face matching. We will apply the IV image overlay system to clinical implementation in the near future.

References

1. F. Sauer, A. Khamene, B. Bascle, and G. J. Rubino, "A head-mounted display system for augmented reality image guidance: Towards clinical evaluation for iMRI-guided neurosurgery," MICCAI 2001, LNCS 2208, pp.707-716, 2001.
2. W. Birkfellner, et al., "A head-mounted operating binocular for augmented reality visualization in medicine - Design and initial evaluation," IEEE Trans. Med. Imag., vol.21, pp.991-997, Aug. 2002.
3. H. Fuchs, et al., "Augmented Reality Visualization for Laparoscopic Surgery," MICCAI 1998, LNCS 1496, pp.934-943, 1998.
4. M. Blackwell, C. Nikou, A. M. Digioia, T. Kanade, "An Image Overlay system for medical data visualization," Medical Image Analysis, Vol.4, pp.67-72, 2000.
5. H. Liao, et al., "Surgical Navigation by Autostereoscopic Image Overlay of Integral Videography," IEEE Trans. Inform. Technol. Biomed., Vol.8 No.2, pp.114-121, June 2004.
6. S. Nicolau, X. Pennec, L. Soler, and N. Ayache, "An Accuracy Certified Augmented Reality System for Therapy Guidance," ECCV 2004, LNCS 3023, pp.79-91, 2004.
7. P. Viola, W. M. Wells III, "Alignment by maximization of mutual information," International Conference on Computer Vision, pp. 16-23,1995.
8. F. M. A. Collignon, et al., "Multimodality Image Registration by Maximization of Mutual Information," IEEE Trans. Med. Imag., Vol.16, pp.187-198, Apr. 1997.
9. J. M. Fitzpatrick, J. B. West, C. R. Maurer Jr., "Predicting Error in Rigid-Body Point-Based Registration," IEEE Trans. Med. Imag., Vol.17, pp.694-702, Oct. 1998.
10. J. P. Wadley, et al., "Neuronavigation in 210 cases: Further development of applications and full integration into contemporary neurosurgical practice," Proc. Computer-Assisted Radiology and Surgery, pp. 635-640, 1998.

Clinical Experience and Perception in Stereo Augmented Reality Surgical Navigation

Philip J. Edwards[1]*, Laura G. Johnson[2], David J. Hawkes[2], Michael R. Fenlon[3], Anthony J. Strong[4], and Michael J. Gleeson[5]

[1] Department of Surgical Oncology and Technology, ICL, 10th Floor QEQM Wing, St. Mary's Hospital, London W2 1NY, UK
Eddie.Edwards@imperial.ac.uk, tel: +44 20 7886 3761
[2] Computational Imaging Science Group (CISG), Imaging Sciences, KCL, Guy's Hospital, London, UK
[3] Dept. of Prosthetic Dentistry., Guy's Hospital, London, UK
[4] Neurosurgery Dept., King's College Hospital, London, UK
[5] ENT surgery Dept., Guy's Hospital, London, UK

Abstract. Over the last 10 years only a small number of systems that provide stereo augmented reality for surgical guidance have been proposed and it has been rare for such devices to be evaluated in-theatre. Over the same period we have developed a system for microscope-assisted guided interventions (MAGI). This provides stereo augmented reality navigation for ENT and neurosurgery. The aim is to enable the surgeon to see virtual structures beneath the operative surface as though the tissue were transparent. During development the system MAGI has undergone regular evaluation in the operating room. This experience has provided valuable feedback for system development as well as insight into the clinical effectiveness of AR. As a result of early difficulties encountered, a parallel project was set up in a laboratory setting to establish the accuracy of depth perception in stereo AR. An interesting anomaly was found when the virtual object is placed 0-40mm below the real surface. Despite this such errors are small (~3mm) and the intraoperative evaluation has established that AR surgical guidance can be clinically effective. Of the 17 cases described here confidence was increased in 3 cases and in a further 2 operations patient outcome was judged to have been improved. Additionally, intuitive and easy craniotomy navigation was achieved even in two cases where registration errors were quite high. These results suggest that stereo AR should have a role in the future of surgical navigation.

Keywords: Image-guided surgery, augmented reality, stereo depth perception

1 Introduction

Augmented reality (AR) surgery guidance aims to combine a real view of the patient on the operating table with virtual renderings of structures that are not

*Corresponding author

G.-Z. Yang and T. Jiang (Eds.): MIAR 2004, LNCS 3150, pp. 369–376, 2004.

visible to the surgeon. A number of systems have been developed that provide such a combined view, usually by overlaying graphics on to video from a device such as an endoscope. If images can be presented in stereo, however, i.e. with a different view to each eye, there is the potential for the position of the real and virtual structures to be seen by the clinician in 3-D.

A handful of stereo AR devices have been proposed for this purpose. Peuchot et al devised a system for spinal surgery in which two half-silvered mirrors were fixed over the patient and from a given viewpoint the surgeon could see a stereo image generated by two overhead displays [1]. Fuchs et al have described a see-through head-mounted display (HMD) that can provide stereo overlays onto a real view of the patient, with proposed applications in ultrasound-guided breast biopsy and 3D laparoscopy [2]. Birkfellner et al have developed the varioscope AR, which provides stereo overlays within a head-mounted microscope system [3]. An entirely video-based stereo HMD AR system has also been proposed by Wendt et al [4].

The HMD systems have the principal disadvantage that any lag between head movements and update of the displays can cause misregistration. As a result these devices have largely been restricted to laboratory-based demonstrations, though work is ongoing to improve their performance. Currently, we are not aware of any stereo AR surgical guidance system having undergone significant clinical evaluation.

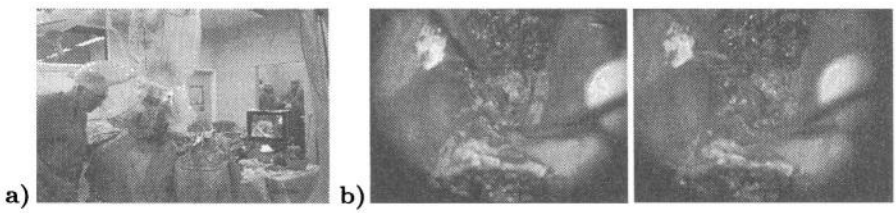


Fig. 1. The MAGI system in-theatre (a) and an example stereo overlay (b), set up for wide-eyed free fusion. The blue vessels should appear to be beneath the surface.

By contrast the MAGI system has been tested in the operating room at regular intervals during its development. The details of the system, which can be seen in figure 1, have been described previously [5,6]. Early results showed that both visualisation and alignment accuracy were significant issues. For visualisation we began a new laboratory project to assess the accuracy of visual perception in stereo AR [7,8]. At the same time much work was done to improve accuracy. For calibration an accurate automated method was developed and for registration we introduced the locking acrylic dental stent (LADS), which attaches to the upper teeth [9]. During these developments the clinical accuracy of registration was assessed by the surgeons as well as the effectiveness of the overall system.

The aim of this paper is to present the results of both the visualisation experiments and the clinical evaluation of MAGI and to suggest future directions for the development of stereo AR surgical guidance.

2 Methods

2.1 Clinical Evaluation

The MAGI project has developed as a collaboration between scientists and clinicians. There has been no attempt at a clinical trial, since the system has changed continually throughout the project as a result of the feedback from operations. The evaluation presented here includes the early development stages as well as the more stable latest version of MAGI.

For each operation the surgeons were asked to assess the performance of MAGI in terms of overlay accuracy, depth perception and clinical effectiveness. Though this is a subjective assessment it gives an evaluation of the utility of the system for the clinical application. We also documented errors which caused the system to fail or lose accuracy.

2.2 Depth Perception Evaluation

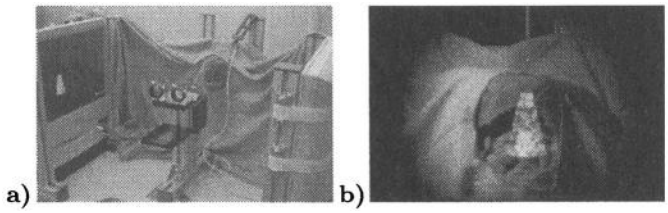


Fig. 2. The MARS laboratory setup (a) and an example overlay of a truncated cone (b).

In order to assess the accuracy of depth perception a laboratory setup known as MARS was developed (see figure 2). This consists simply of two LCD monitors and two beamsplitters with eyepieces to constrain the viewing position. This avoids any of the optics involved with the stereo microscope that may influence perception. A new calibration technique was developed that calculates the position of the virtual image of each monitor. An individual calibration is then performed in which the pointer is marked at several positions away from this virtual image to calculate the position of each observers pupil. This compensates for any interpupillary distance variation for the different subjects.

A truncated cone is overlaid using MARS on a phantom that mimics skin and brain. The cone can be placed at various depths beneath (or above) the viewed physical surface. The subject is then asked to mark the position of the tip of

the cone, which is outside the phantom, with a tracked pointer. In this way the depth of an object entirely within the phantom can be marked (see figure 2(b)).

In one experiment there was simply a comparison of the accuracy with and without the phantom present. This shows the effect the presence of a visible surface in front of the virtual cone has on perception accuracy. There were 8 observers and their results were pooled from 60 measurements of the cone at varying depths below the real surface.

A further experiment examined the profile of perception error with depth beneath the surface of the phantom. The cone was displayed at 5mm intervals from 20mm in front of the phantom to 20mm behind and also at 40mm and 80mm. The measurements were pooled across 5 observers, who marked each position 12 times, by first subtracting the average error for the four positions above the surface to reduce any residual calibration error. In this way we examine only the variation of error with depth.

3 Results

3.1 Clinical Evaluation Results

The results of the clinical evaluation can be seen in table 1. To summarise, of the 17 operations the accuracy was judged to be 1mm or better in 5 cases (29%), 2mm or better in 11 cases (65%).

The operations where errors were greater are worth considering in some detail. There were 3 cases in which errors of greater than 3mm were recorded. For patient A, early in the project development, anatomical landmark registration was used which we know is prone to inaccuracy. This led to the development of other registration strategies. For patient F, soft tissue movement, or brain shift, had occurred. This is known to be a problem for neurosurgical guidance [10,11]. In the case of patient J the LADS was a poor fit due to the fact that few teeth were available. This underlines the need for careful patient selection when the LADS is to be used.

Despite high errors, in two of these cases craniotomy planning was possible. For patients F and J the craniotomy site was adjusted using the system and found to be well positioned. Stereo AR guidance enables craniotomy planning to be carried out in a simple and intuitive manner since the position of the lesion can be visualised directly on the patient.

Registration was judged to have failed in 2 cases. Patient B was our first attempt at an alternative registration technique. Here, markers were attached to a radiotherapy mask in the hope that these would provide a more accurate registration than anatomical landmarks. Though face moulds are an accurate technique for radiotherapy, where imaging and therapy are performed with the patient in a similar position, skin motion causes significant errors for use during surgery. It was after this procedure that we began development of the LADS. In the case of patient N draping was performed in a manner that did not leave sufficient line-of-sight to the tracking markers on the LADS. This demonstrates the

Table 1. List of operations and MAGI performance evaluation.

Patient ID	Pathology	Registration Accuracy (mm)	Effect on Procedure	Comments
A	Epidermoid cyst	3-5	None	Anatomical Registration
B	Glomus jugulare	>10 (fail)	None	Radiotherapy mask registration
C	Ethmoid carcinoma	2 (initial)	None	Initial LADS design (unstable)
D	Carotid lesion	2-3	None	LADS Mk II (heavy)
E	Petrous apex cyst	1	Increased confidence	Current LADS design
F	Superficial AVM	3-4	Craniotomy navigation	Brain shift
G	Facial nerve angioma	<1	None	Accurate LADS
H	Clivus meningioma	1-2	Uncomfortable bone-pin position	-
I	Vestibular Schwannoma	2	None	Good 3D perception (moving texture)
J	Parietal meningioma	~3-10	Craniotomy navigation	Poor LADS fit (few teeth)
K	Vestibular Schwannoma (NF2)	1-2	Improved outcome	Aided in long procedure
L	Subfrontal meningioma	1-2	None	Excellent 3D perception of vessels
M	Olfactory neuroblastoma	<1	None	Accurate LADS repositioning
N	Subfrontal meningioma	fail	none	Inappropriate draping
M	Recurrent ethmoid meningioma	1	Full extraction	Pointer only with LADS
N	Ethmoid carcinoma	1-2	Improved confidence	Imaging 1 week earlier + pins
O	Adenocarcinoma	1	improved confidence	Bone surface registration

importance of having someone who knows the guidance system present during this phase of the operation.

Apart from the craniotomy examples, the system was found to be beneficial in a number of other cases. Patient E had bilateral petrous apex cysts with only one side being symptomatic. Due to the possibility of the need for a further operation on the contralateral side it was decided to take an alternative path to the more common translabyrinthine approach to preserve hearing. This required high accuracy to avoid damage to either the carotid artery or facial nerve. The operation was successful and the surgeon commented that MAGI has been a significant aid to confidence in an unfamiliar approach.

Patient N had an ethmoid carcinoma extending into the left orbit and was also blind in the right eye. The aim of surgery was to remove the lesion with as little damage to the optic nerve as possible. It was reported that the system again improved the surgeon's confidence, in particular that the correct position had been reached along the ethmoid without invading the sphenoid sinus.

Patient O had a low grade recurrent adenocarcinoma of the ethmoid not unlike patient N. A similar approach was taken and the system was judged to have aided confidence in a similar manner, also aiding in achieving full extraction of the lesion.

With patient M there was insufficient time between imaging and the operation to perform segmentation for overlays. Nonetheless the pointer-based guidance proved highly valuable and showed the accuracy of registration achieved with the LADS. This patient had abnormal anatomy that meant that a portion of the lesion, a meningioma again in the ethmoid, would not have been extracted without guidance.

Patient K had a very large vestibular Schwannoma as a result of neurofibromatosis 2. This was a potentially very hazardous procedure as the lesion was close to several critical structures including the brain stem and basilar artery and the approach could involve the carotid artery, lateral sinus or jugular. One ENT surgeon and two neurosurgeons participated in an operation that lasted some 15 hours. At several points the system was used while surgery proceeded in a manner that would not be possible with pointer-based guidance. All three surgeons reported that the system had improved both confidence and outcome for the patient. Though full extraction was not achieved there was no significant damage to vital structures and the patient showed great improvement as a result of surgery.

3.2 Depth Perception Results

In the first experiment, the error magnitude in marking the virtual cone without a real surface present was 1.05±0.25mm. With the physical surface present this rises to 2.1±0.3mm. This demonstrates that the presence of the physical surface has an effect on accuracy, but does not give any information about the direction of the error or its relationship to the depth of the cone.

The results of the second experiment are shown in figure 3. All subjects showed a tendency to see the virtual cone as deeper into the surface than its

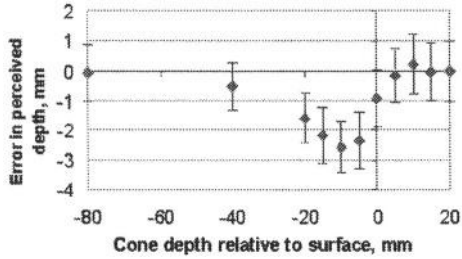


Fig. 3. Profile of error in depth perception, showing that the virtual cone is perceived as deeper than its actual position when less than 20mm deep to the real surface.

true position for depths of 0-20mm. The peak error appears at approximately 15mm below the surface and there is a characteristic shape to the graph.

4 Conclusions

This paper presents the first clinical evaluation of a stereo augmented reality system for surgery guidance. Though there were 5 failed or inaccurate registrations these were due to specific mistakes that led to the development of new methodology or changes in practice. It is hoped that inclusion of these cases in the paper will provide useful information for those developing such systems.

Of the remaining cases the average accuracy was ~1.5mm, which is a good accuracy for any guidance system but particularly impressive for an AR system. Where confidence is improved, as happened in three cases, there is the possibility that the system will make a difference to outcome or perhaps decrease the time taken for an operation. It should be borne in mind that overconfidence in an inaccurate system could potentially be damaging, but this was not the case in any of our procedures. In two cases it was assessed that clinical outcome had definitely been improved by MAGI.

Early problems with 3D perception led to the development of the MARS system. The presence of a physical surface causes the virtual surface to be perceived up to 3mm deeper than its displayed position with a characteristic shape to the dependence on depth. This could potentially be a problem if a critical structure is assumed to be further than its true position, though the error should reduce as the surgeon works towards the target. Though this phenomenon has been established in the laboratory it has not proven to be a problem in any of our clinical cases with MAGI so far. Also we are investigating whether other graphical cues can reduce or eliminate this error.

It is hoped that this account will encourage those who are developing AR navigation systems to involve surgeons and to evaluate their devices in the clinical environment at as early a stage as possible. We believe we have already demonstrated the potential of AR as a means of guiding surgery and hope that this work strengthens the case for its continued development.

Acknowledgements

This project has been funded at various points by Leica, BrainLAB and EPSRC and we are grateful for their support. We would also like to thank the radiology, radiography and theatre staff at Guy's and King's hospitals for their cooperation. Also thanks to the research students and staff at CISG who have contributed to MAGI, namely Bassem Ismail, Oliver Fleig and in particular Dr. Andy King.

References

1. Peuchot, B., Tanguy, A., Eude, M.: Virtual reality as an operative tool during scoliosis surgery. In Ayache, N., ed.: Computer Vision, Virtual Reality and Robotics in Medicine. Lecture Notes in Computer Science (905), Springer-Verlag (1995) 549–554
2. Fuchs, H., Livingston, M.A., Raskar, R., Colucci, D., Keller, K., State, A., Crawford, J.R., Rademacher, P., Drake, S.H., Meyer, A.A.: Augmented reality visualization for laparoscopic surgery. In: Proc. Medical Image Computation and Computer-Assisted Intervention. (1998) 934–943
3. Birkfellner, W., Figl, M., Huber, K., Watzinger, F., Wanschitz, F., Hummel, J., Hanel, R., Greimel, W., Homolka, P., Ewers, R., Bergmann, H.: A head-mounted operating binocular for augmented reality visualization in medicine - design and initial evaluation. IEEE Trans. Med. Imaging **21** (2002) 991–997
4. Wendt, M., Sauer, F., Khamene, A., Bascle, B., Vogt, S., Wacker, F.K.: A head-mounted display system for augmented reality: Initial evaluation for interventional mri. Rofo-Fortschr. Gebiet Rontgenstrahlen Bildgeb. Verfahr. **175** (2003) 418–421
5. Edwards, P.J., Hawkes, D.J., Hill, D.L.G., Jewell, D., Spink, R., Strong, A.J., Gleeson, M.J.: Augmentation of reality in the stereo operating microscope for otolaryngology and neurosurgical guidance. J. Image Guid. Surg. **1** (1995) 172–178
6. Edwards, P.J., King, A.P., Maurer, Jr., C.R., de Cunha, D.A., Hawkes, D.J., Hill, D.L.G., Gaston, R.P., Fenlon, M.R., Jusczyzck, A., Strong, A.J., Chandler, C.L., Gleeson, M.J.: Design and evaluation of a system for microscope-assisted guided interventions (MAGI). IEEE Trans. Med. Imaging **19** (2000) 1082–1093
7. Johnson, L.G., Edwards, P.J., Hawkes, D.J.: Surface transparency makes stereo overlays unpredictable: The implications for augmented reality. In: Medicine Meets Virtual Reality. Health Technology and Informatics, IOS Press (2003) 131–136
8. Johnson, L.G., Edwards, P.J., Barratt, D.C., Hawkes, D.J.: The problem of perceiving accurate depth information through a surgical augmented reality system. In: Proc. Computer Assisted Surgery around the Head (CAS-H 2003). (2003)
9. Fenlon, M.R., Jusczyzck, A.S., Edwards, P.J., King, A.P.: Locking acrylic resin dental stent for image-guided surgery. J. Prosthet. Dent. **83** (2000) 482–485
10. Roberts, D.W., Hartov, A., Kennedy, F.E., Miga, M.I., Paulsen, K.D.: Intraoperative brain shift and deformation: A quantitative analysis of cortical displacement in 28 cases. Neurosurgery **43** (1998) 749–760
11. Hill, D.L.G., Maurer, Jr., C.R., Maciunas, R.J., Barwise, J. A. Fitzpatrick, J.M., Wang, M.Y.: Measurement of intraoperative brain surface deformation under a craniotomy. Neurosurgery **43** (1998) 514–528

Author Index