Hongen Liao P.J. "Eddie" Edwards
Xiaochuan Pan Yong Fan
Guang-Zhong Yang (Eds.)

LNCS 6326

Medical Imaging and Augmented Reality

5th International Workshop, MIAR 2010
Beijing, China, September 2010
Proceedings

Springer

Lecture Notes in Computer Science 6326

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

Hongen Liao P.J. "Eddie" Edwards
Xiaochuan Pan Yong Fan
Guang-Zhong Yang (Eds.)

Medical Imaging and Augmented Reality

5th International Workshop, MIAR 2010
Beijing, China, September 19-20, 2010
Proceedings

Springer

Volume Editors

Hongen Liao
The University of Tokyo, Department of Bioengineering
7-3-1, Hongo, Bunkyo, Tokyo, 113-8656, Japan
E-mail: liao@bmpe.t.u-tokyo.ac.jp

P.J. "Eddie" Edwards
Imperial College London, Department of Surgery and Cancer
South Kensington Campus, London SW7 2AZ, UK
E-mail: eddie.edwards@imperial.ac.uk

Xiaochuan Pan
The University of Chicago, Department of Radiology
5841 South Maryland Avenue, Chicago, IL 60637, USA
E-mail: xpan@uchicago.edu

Yong Fan
Chinese Academy of Sciences, Institute of Automation
P.O. Box 2728, Beijing 100080, China
E-mail: yfan@nlpr.ia.ac.cn

Guang-Zhong Yang
Imperial College London, Institute of Biomedical Engineering
180 Queens Gate, London SW7 2BZ, UK
E-mail: g.z.yang@imperial.ac.uk

Library of Congress Control Number: Applied for

CR Subject Classification (1998): J.3, I.4, I.6, I.5, I.2.10, I.3.5

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

ISSN 0302-9743
ISBN-10 3-642-15698-3 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-15698-4 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper 06/3180

Preface

The 5th International Workshop on Medical Imaging and Augmented Reality, MIAR 2010, was held at the China National Convention Center (CNCC), Beijing, China on September 19–20, 2010.

MIAR has remained a truly international meeting, bringing together researchers from all fields related to medical image analysis, visualization and targeted intervention. In recent years, technical advances in therapeutic delivery and a growing demand for patient-specific treatment have accelerated the clinical applications of MIAR-related techniques. Imaging plays an increasingly important role in targeted therapy, with interventions such as drug or gene therapy relying on more accurate delivery tailored to individual patients. Rapid progress in surgical methodologies, such as those with robot assistance, demands precise guidance from both preoperative and intraoperative imaging. The volume of data available from existing and emerging imaging modalities leads to a desire for more automated analysis for diagnosis, segmentation and registration. Research in this rapidly developing area is highly multi-disciplinary, integrating research in life sciences, physical sciences, engineering, and medicine.

As a high impact workshop, MIAR continues to grow. For this year, we received 139 full papers covering medical image formation, analysis and interpretation; augmented reality, visualization and simulation; computer assisted intervention and robotics; surgical planning; systematic extra- and intra-corporeal imaging modalities; general biological and neuroscience image computing; and patient specific modeling and medical image understanding. The papers were judged by up to five reviewers in a double-blind review process. The quality of the submissions was excellent and 60 papers (43%) were accepted by the Program Committee. These papers were presented in a single track of oral and poster sessions.

The organization of MIAR was very much a team effort and we are extremely grateful to all members of the Program Committee and also members of the International and Local Organizing Committees. The review process was kept to a very tight schedule and we appreciate the commitment and professionalism shown by all those who took part.

We also thank the invited speaker Prof. Zhi-Pei Liang from the University of Illinois at Urbana-Champaign, USA, for his lecture on fast imaging with sparse sampling. We are also grateful to Stephane Nicolau from IRCAD Taiwan, for his invited talk on augmented reality surgical guidance.

We were delighted to host MIAR in Beijing, one of the world's leading business, cultural, and political centers. Its influence on education, entertainment, media, fashion, and the Arts also contributes to its status as a major global city. CNCC is the largest and newest international conference center in China. Situated in the heart of Olympic Green, CNCC is right next to the Bird's Nest

(National Stadium) and the Water Cube (National Aquatics Center) and enjoys unparalleled, easy access to all parts of this vibrant city.

For those who were not able to join us at MIAR 2010, we hope this volume will serve as a valuable reference and we hope to see you at future MIAR workshops.

September 2010

Hongen Liao
PJ "Eddie" Edwards
Xiaochuan Pan
Yong Fan
Guang-Zhong Yang

Organization

Executive Committee

General Chair and Co-chair

Hongen Liao (Chair)	The University of Tokyo, Japan
PJ "Eddie" Edwards	Imperial College London, UK

Program Chair and Co-chairs

Xiaochuan Pan (Chair)	University of Chicago, USA
Yong Fan	Chinese Academy of Sciences, China
Guang-Zhong Yang	Imperial College London, UK

Program Committee

Leon Axel	NYU Medical Center, USA
Shanglian Bao	Peking University, China
Christian Barillot	IRISA, Rennes, France
Adrien Bartoli	LASMEA, Clermont-Ferrand, France
Wolfgang Birkfellner	Medical University of Vienna, Austria
Yiyu Cai	Nanyang Technological University, Singapore
Yen-wei Chen	Ritsumeikan University, Japan
Takeyoshi Dohi	The University of Tokyo, Japan
Gary Egan	Howard Florey Institute, Australia
Jing Fang	Peking University, China
Gabor Fichtinger	Queen's University, Canada
James Gee	University of Pennsylvania, USA
Guido Gerig	University of North Carolina at Chapel Hill, USA
Lixu Gu	Shanghai Jiaotong University, China
Makoto Hashizume	Kyushu University, Japan
David J. Hawkes	University College London, UK
Pierre Jannin	IRISA, France
Ron Kikinis	Harvard Medical School, USA
Young-Soo Kim	Hanyang University, South Korea
Tianming Liu	University of Georgia, USA
Jimmy Liu	Agency for Science, Technology and Research, Singapore
Anthony Maeder	CSIRO, Australia
Ken Masamune	The University of Tokyo, Japan
Kensaku Mori	Nagoya University, Japan

Nassir Navab	Technical University of Munich, Germany
Terry Peters	Robarts Research Institute, Canada
Stephen Riederer	Mayo Clinic, USA
Ichiro Sakuma	The University of Tokyo, Japan
Yoshinobu Sato	Osaka University, Japan
Dinggang Shen	UNC at Chapel Hill, USA
Pengcheng Shi	Rochester Institute of Technology, USA
Russell H. Taylor	Johns Hopkins University, USA
Guangzhi Wang	Tsinghua University, China
Stephen T.C. Wong	Methodist Hospital - Weill Cornell Medical College, USA

International Organizing Committee

Dirk Bartz (†)	University of Leipzig, Germany
Chung-Ming Chen	National Taiwan University, Taiwan
Elvis Chen	Robarts Research Institute, Canada
Adrian Chung	Imperial College London, UK
Hui Ding	Tsinghua University, China
Michael Figl	Medical University of Vienna, Austria
Ali Gooya	University of Pennsylvania, USA
Jaesung Hong	Kyushu University, Japan
Ali Khamene	Siemens Corp. Research, USA
Jan Klein	Fraunhofer MEVIS, Germany
Su-Lin Lee	Imperial College London, UK
Mirna Lerotic	Imperial College London, UK
Ming Li	National Institutes of Health, USA
Pai-Chi Li	National Taiwan University, Taiwan
John Moore	Robarts Research Institute, Canada
Ryoichi Nakamura	Chiba University, Japan
Kilian Pohl	IBM Research, USA
Jaw-Lin Wang	National Taiwan University, Taiwan
Kelvin K. Wong	Methodist Hospital - Weill Cornell Medical College, USA
Yaoqin Xie	Chinese Academy of Sciences, China
Zhong Xue	Weill Cornell Medical College, USA
Jong Chul Ye	Korea Advanced Institute of Science and Technology, South Korea
Jue Zhang	Peking University, China
Bo Zheng	The University of Tokyo, Japan
Guoyan Zheng	University of Bern, Switzerland

Local Organizing Committee

Ming Song	Chinese Academy of Sciences, China
Yong Liu	Chinese Academy of Sciences, China
Bing Liu	Chinese Academy of Sciences, China
Yonghui Li	Chinese Academy of Sciences, China
Shaomei Wang	Chinese Academy of Sciences, China
Nianming Zuo	Chinese Academy of Sciences, China

Sponsors and Sponsoring Institutions

Chinese Academy of Sciences
NDI-Northern Digital Inc.
Siemens Corporate Research
Translational Systems Biology and Medicine Initiative
The University of Tokyo

Reviewers

Axel, Leon
Baloch, Sajjad
Barillot, Christian
Bartoli, Adrien
Cai, Yiyu
Chen, Elvis C.S.
Chen, Yasheng
Chen, Yen-Wei
Cheng, Jie
Chung, Adrian
Commowick, Olivier
Deng, Xiang
Ding, Hui
Drechsler, Klaus
Edwards, P.J. "Eddie"
Essert, Caroline
Fan, Yong
Fichtinger, Gabor
Figl, Michael
Gee, James
Gerig, Guido
Giannarou, Stamatia
Gooya, Ali
Greenleef, James
Gu, Lixu
Guo, Yujun
Hammoudi, Ahmad
Hashizume, Makoto
Ho, Harvey
Hong, Jaesung
Hu, Liang
Hu, Yipeng
Huang, Min
Jain, Ameet
Jannin, Pierre
Jia, Hongjun
Jiao, Fangxiang
Joshi, Alark
Kersten, Marta
Kikinis, Ron
Kim, Young-Soo
Klein, Jan
Kuroda, Yoshihiro
Lalys, Florent
Lapeer, Rudy
Lee, Su-Lin
Lefevre, Julien
Lerotic, Mirna
Li, Fuhai
Li, Gang
Li, Ming
Liao, Hongen

Liao, Rui
Lin, Shane
Linte, Cristian
Liu, Jimmy
Liu, Tianming
Liu, Xiaofeng
Lu, Kongkuo
Luan, Kuan
Masamune, Ken
Mercier, Laurence
Milgram, Paul
Moore, John
Mori, Kensaku
Nakamura, Ryoichi
Ogiela, Marek
Pan, Xiaochuan
Peters, Terry
Pohl, Kilian
Qazi, Arish
Riederer, Stephen J.
Ruan, Su
Sato, Yoshinobu
Satoh, Kiyohide
Shen, Dinggang
Shen, Li
Shi, Feng
Shi, Pengcheng
Stoyanov, Danail
Tao, Xiaodong
Taylor, Russell H.
Totz, Johannes
Traub, Joerg
Tustison, Nicholas
Vrtovec, Tomaz
Wang, Guangzhi
Wang, Junchen
Wang, Li
Wang, Qian
Wang, Zhiyong
Ward, Aaron
Wesarg, Stefan
Wiles, Andrew
Wong, Kelvin K.
Wong, Stephen TC
Wu, Guorong
Xie, Yaoqin
Xue, Zhong
Yamauchi, Yasushi
Yan, Jianhua
Yang, Guang-Zhong
Yang, Li
Yang, Yaliang
Yang, Zhengyi
Yap, Pew-Thian
Zhang, Shaoting
Zheng, Bo
Zheng, Guoyan
Zhuang, Xiahai

Table of Contents

Image Segmentation

Image Registration

Shape Modeling and Morphometry

Image Analysis

Diffiusion Tensor Image

Computer Assisted Intervention

Medical Image Computing

Visualization and Application

Segmentation and Classification

Medical Image Understanding

Image-Guided Surgery

Augmented Reality

Automatic Segmentation of Neonatal Images Using Convex Optimization and Coupled Level Set Method

Li Wang[1,2], Feng Shi[2], John H. Gilmore[3], Weili Lin[4], and Dinggang Shen[2,⋆]

[1]School of Computer Science & Technology,
Nanjing University of Science and Technology, China
[2]IDEA Lab, Department of Radiology and BRIC,
University of North Carolina at Chapel Hill, USA
dgshen@med.unc.edu
[3]Department of Psychiatry, University of North Carolina at Chapel Hill, USA
[4]MRI Lab, Department of Radiology and BRIC,
University of North Carolina at Chapel Hill, USA

Abstract. Accurate segmentation of neonatal brain MR images remains challenging mainly due to poor spatial resolution, low tissue contrast, high intensity inhomogeneity. Most existing methods for neonatal brain segmentation are atlas-based and voxel-wise. Although parametric or geometric deformable models have been successfully applied to adult brain segmentation, to the best of our knowledge, they are not explored in neonatal images. In this paper, we propose a novel neonatal image segmentation method, combining local intensity information, atlas spatial prior and cortical thickness constraint, in a level set framework. Besides, we also provide a robust and reliable tissue surfaces initialization for our proposed level set method by using a convex optimization technique. Validation is performed on 10 neonatal brain images with promising results.

1 Introduction

Accurate segmentation of neonatal brain structures from magnetic resonance (MR) images has important implications for normal brain development, as well as for the diagnose and treatment of neurodevelepmental disorders such as prematurity. Manual segmentation of neonatal brain structures is tedious, time consuming, and also lacks of reproducibility. Therefore, it is necessary to use automatic techniques for neonatal brain segmentation. However, despite of the success of segmentation methods developed for adult brain, it still remains challenging to segment neonatal brain images [1,2] due to poor spatial resolution, low contrast, and ambiguous tissue intensity distribution [1,3], as well as the inverted contrast between white matter (WM) and gray matter (GM) [2].

Most existing methods for neonatal brain segmentation are atlas-based and voxel-wise [1,3,4,5,6]. For example, Prastawa *et al.* [1] proposed an atlas-based approach for neonatal brain segmentation. They generated an atlas by averaging three semi-automatic segmented neonatal brain images and adopted the

⋆ Corresponding author.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 1–10, 2010.

expectation-maximization (EM) scheme with inhomogeneity correction to achieve tissue classification. Shi *et al.* [3] proposed a framework for performing neonatal brain tissue segmentation by using a subject-specific tissue probabilistic atlas generated from longitudinal data follow-up of the same subject. All the above-mentioned methods for neonatal segmentation, however, are based on voxel-wise segmentation. Geometric information has not been paid much attention in the neonatal brain segmentation. However, geometric information describes the gradient and boundary of tissue structures, constraints the relationship of structural shapes, which is appreciated in tissue segmentation to manage the ambiguous structural distributions, especially in neonatal images.

One of the most effective ways of incorporating geometric information for tissue segmentation is to use active contour/surface models [7]. These models are able to provide smooth and closed contours/surfaces as final segmentation, which is not possible for the voxel-based segmentation methods. In fact, geometrically, the human cerebral cortex is a thin, folded sheet of GM, with a nearly consistent thickness of 1-5 mm for neonatal brains. Therefore, surface-based techniques are considered to be more suitable for neonatal brain segmentation than the voxel-based segmentation methods. To obtain a detailed geometric representation of the cortex, many algorithms have been proposed using explicit or implicit surface representation [8,9,10,11,12]. However, they cannot be directly applied to neonatal brain images.

2 Method

In this paper, we present a novel surface-based method, utilizing local intensity information, atlas spatial prior and cortical thickness constraint, for segmentation of neonatal MR brain images. We adopt the local Gaussian distribution fitting (LGDF) energy [13], which describes local image intensities by Gaussian distributions with different means and variances. The means and variances of local intensities are spatially varying functions, which enable the model to deal with intensity inhomogeneities. A prior knowledge from atlases is then combined with the LGDF energy to regularize the segmentation and further increase its ability of handling inhomogeneities. Based on the fact that the cortex has a nearly constant thickness, a constraint of cortical thickness can provide useful geometric information to guide more accurate segmentation. Accordingly, these three terms are finally incorporated into a coupled surface-based method in such a way that the surfaces are driven by the LGDF and spatial prior, while the distance between the inner and the outer surfaces of cortex remains within a predefined range by the constraint of cortical thickness. The contributions of this paper are three-fold:

a) We use adaptive mean and variance for the same tissue at different locations of brain, for dealing with the inhomogeneities;
b) We use atlas prior information to guide the segmentation;
c) We use coupled surfaces to fit the CSF/GM and GM/WM boundaries with a cortical thickness constraint.

An overview of the proposed framework is shown in Fig. 1. The framework consists of three steps: (1) Preliminary segmentation for CSF, WM and GM, as shown in the right panel of Fig. 1; (2) Partial Volume (PV) removal and correction of the mislabeled CSF from WM, as shown in the bottom panel; and (3) Coupled surface-based segmentation, as shown in the left panel. Steps (1) and (2) form an initialization for the step (3). For better emphasizing our contribution, we will first introduce step (3) in Section 2.1, and then steps (1) and (2) in Sections 2.2 and 2.3, respectively. The following sections describe the method in detail.

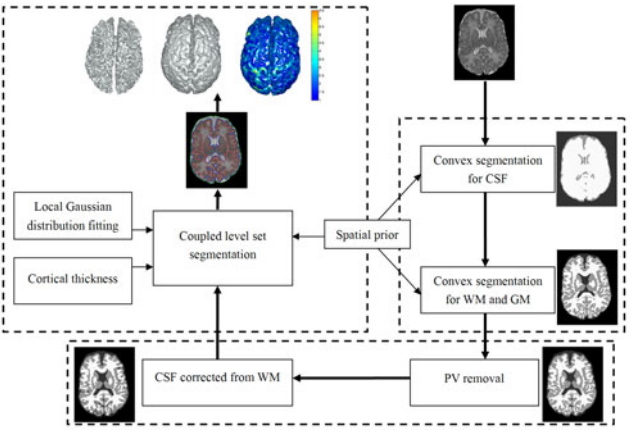


Fig. 1. The proposed framework for neonatal segmentation

2.1 Neonatal segmentation Using Coupled Level Set Method

In this section, we propose an implicit level set method based on local intensity distribution fitting, spatial prior, and cortical thickness constraint for neonatal brain segmentation. Let Ω be the image domain, I be a given image, and $\{\Omega_i\}_{i=1}^N$ be a set of disjoint image regions, such that $\Omega = \cup_{i=1}^N \Omega_i$, $\Omega_i \cap \Omega_j = \emptyset$, $\forall i \neq j$, where N refers to the number of regions. Based on the work in [13], for each point x in the image domain Ω, the local Gaussian distribution fitting energy is defined as $\mathcal{E}_x^{LGDF} = \sum_{i=1}^N \int_{\Omega_i} -\omega_\sigma(x-y) \log p_{i,x}(I(y))dy$, where $\omega_\sigma(x-y)$ is a Gaussian kernel with a scale parameter σ as proposed in [14,15] and $p_{i,x}(I(y))$ is the probability density, which is defined as $p_{i,x}(I(y)) = 1/(\sqrt{2\pi}\sigma_i(x)) \exp(-(u_i(x) - I(y))^2/(2\sigma_i(x)^2))$, where $u_i(x)$ and $\sigma_i(x)$ are local intensity means and standard deviations, respectively. It is worth noting that local intensity means $u_i(x)$ and variances $\sigma_i^2(x)$ are the spatially varying functions, which are crucial in handling the inhomogeneity. Due to large overlap in the tissue distribution, it is necessary to use spatial prior for guiding the segmentation. In the following, we propose a new energy function which combines the local Gaussian distribution fitting energy and spatial prior knowledge $prior_i$ from neonatal brain atlases,

$$\mathcal{E}_x^{L-Prior} = \sum_{i=1}^{N} \int_{\Omega_i} -\omega_\sigma(x-y) \log(prior_i(y) p_{i,x}(I(y)))dy \quad (1)$$

The ultimate goal is to minimize $\mathcal{E}_x^{L-Prior}$ for all the center points x in the image domain Ω, which directs us to define an energy function as the following double integral: $\mathcal{E}^{L-Prior} = \int \mathcal{E}_x^{L-Prior} dx$. We can use one or multiple level set functions to represent a partition $\{\Omega_i\}_{i=1}^N$. For neonatal segmentation, we use three level set functions ϕ_1, ϕ_2 and ϕ_3 to represent WM, GM, CSF and background, where the zero level surfaces of ϕ_1, ϕ_2 and ϕ_3 are interfaces of WM/GM, GM/CSF, and CSF/background, respectively. Let $\Phi = (\phi_1, \phi_2, \phi_3)$. Using Heaviside function H, the energy function based on the LGDF energy and atlas spatial prior can be defined as

$$\mathcal{F} = \int \left(\sum_{i=1}^{4} \int -\omega_\sigma(x-y) \log(prior_i(y) p_{i,x}(I(y))) M_i(\Phi(y)) dy \right) dx + \nu \sum_{i=1}^{3} \mathcal{L}(\phi_i) \quad (2)$$

where $\mathcal{L}(\phi_i) = \int |\nabla H(\phi_i(x))| dx$ is the length term to maintain a smooth contour/surface during evolution, and $M_i(\Phi)$ are defined as $M_1 = H(\phi_1)H(\phi_2)H(\phi_3)$, $M_2 = (1-H(\phi_1))H(\phi_2)H(\phi_3)$, $M_3 = (1-H(\phi_2))H(\phi_3)$, and $M_4 = 1-H(\phi_3)$.

Minimization of the energy function $\mathcal{F}$ in Eq. (2) with respect to ϕ_i is achieved by solving the gradient descent flow equations as follows,

$$\begin{aligned}
\partial\phi_1/\partial t &= -\delta(\phi_1)H(\phi_2)(e_1-e_2)H(\phi_3) + \nu\delta(\phi_1)K_1 \\
\partial\phi_2/\partial t &= -\delta(\phi_2)(H(\phi_1)(e_1-e_2) + (e_2-e_3))H(\phi_3) + \nu\delta(\phi_2)K_2 \\
\partial\phi_3/\partial t &= -\delta(\phi_3)(H(\phi_2)H(\phi_1)e_1 + H(\phi_2)(1-H(\phi_1))e_2 + (1-H(\phi_2))e_3 - e_4) + \nu\delta(\phi_3)K_3
\end{aligned} \quad (3)$$

where $K_i = \text{div}\left(\frac{\nabla\phi_i}{|\nabla\phi_i|}\right)$ and $e_i(x) = -log(prior_i(x)) + \int \omega_\sigma(y-x)[\log(\sigma_i(y)) + \frac{(u_i(y)-I(x))^2}{2\sigma_i(y)^2}]dy$.

As proposed in [9,11], the cortex layer has a nearly consistent thickness which can be used to guide the surface evolutions. To utilize the cortical structural information, we design a coupled surfaces model to constrain the distance of zeros level surfaces of ϕ_1 and ϕ_2 within a reasonable range. Let the allowed distance be $[d\ D]$. We adopt the coupling functions $h(\cdot)$ and $c_i(\phi_j)$ in [16,9], where $h(x)$ is a function that $h(x) = 1$ when the distance between the two surfaces is within an acceptable range, otherwise $h(x) = 0$, and $c_i(\phi_j)$ is another coupling function that remains the distance within an acceptable range. Therefore, we write

$$\partial\phi_1/\partial t = h(|\phi_2|)\left[-\delta(\phi_1)H(\phi_2)(e_1-e_2)H(\phi_3)\right] + c_1(\phi_2)|\nabla\phi_1| + \nu\delta(\phi_1)K_1 \quad (4)$$

where $h(\cdot)$ and $c_i(\phi_j)$ are defined as

$$h(x) = \begin{cases} 0, & [x \le d] \cup [x > D]; \\ 1, & [d+f < x < D-f]; \\ 1-(\frac{x-d-f}{f})^2, & [d < x \le d+f]; \\ 1-(\frac{x-D+f}{f})^2, & [D-f \le x \le D]. \end{cases} \quad c_i(\phi_j) = (1-h(|\phi_j|)) \begin{cases} sign(\phi_j), & |\phi_j| \ge D-f; \\ -sign(\phi_j), & |\phi_j| \le d+f. \end{cases}$$

where f is the constant that determines the range of the preferable values for the distance $[d+f, D-f]$ (In all our experiments, we set $f=1$). There are two advantages of this design. First, the image-based force $[-\delta(\phi_1)H(\phi_2)(e_1 - e_2)H(\phi_3)]$

guides the surface evolution only when the distance between the two surfaces is within the acceptable range. Second, when the distance is beyond the acceptable range, this force does not affect the evolutions but the second term $c_1(\phi_2)|\nabla\phi_1|$ is activated which will deflate the surface if the distance is below the minimum acceptable value, and inflate the surface if the distance is beyond the maximum acceptable value.

In a similar way, we write a new evolution equation for ϕ_2,

$$\partial\phi_2/\partial t = h(|\phi_1|)[-\delta(\phi_2)(H(\phi_1)(e_1-e_2)+(e_2-e_3))H(\phi_3)]+c_2(\phi_1)|\nabla\phi_2|+\nu\delta(\phi_2)K_2 \quad (5)$$

With the combination of local Gaussian distribution fitting energy, spatial prior knowledge, and cortical thickness constraint, the proposed method is able to achieve accurate segmentation for neonatal MR images. However, as 3D convolution operations are performed every iteration, the proposed method is computationally expensive. A good initialization for the proposed method is not only necessary to save time but also help avoid being trapped into local minima. In the following section, we will propose a robust initialization method based on convex optimization.

2.2 Preliminary Segmentation for CSF, WM and GM

Due to the fact that CSF has the highest intensity in neonatal T2 brain image, we can first extract CSF from the brain image, and then separate WM from GM. This design is much easier than the extraction of CSF, WM and GM simultaneously as in the conventional methods. To address the issue of inhomogeneity, we adopt a joint segmentation and inhomogeneity estimation scheme. We first logarithmically transform the intensities in order to make the bias b additive. We then use two variables u_1 and u_2, which take values between 0 and 1, to represent the membership functions of three regions with $M_1 = u_1u_2$, $M_2 = u_1(1-u_2)$, and $M_3 = (1-u_1)$. The intensities of each region are characterized by a global Gaussian distribution with mean and variance as (c_i, σ_i^2). The atlas spatial prior $prior_i$ is also utilized for segmentation. We then propose the following energy for segmenting the image into the regions of CSF, (WM+GM) and background,

$$\begin{aligned} E(u_1, u_2, c_i, \sigma_i, b) = &-\textstyle\sum_{i=1}^{3} \int \log(prior_i(x)p_i(x))M_i(x))dx \\ &+ \nu \textstyle\int |\nabla u_1|dx + \nu \int |\nabla u_2|dx, u_1 \in [0,1], u_2 \in [0,1] \end{aligned} \quad (6)$$

where $p_i(x) = 1/(\sqrt{2\pi}\sigma_i)\exp(-(\log I(x) - c_i - b)^2/(2\sigma_i^2))$ and the last two terms are the total variation of u_1 and u_2. By constraining both u_1 and u_2 to be $[0, 1]$, the minimization problem is convex with respect to u_1 and u_2 when (c_i, σ_i, b) are fixed [17,18]. Therefore, we can easily use the Split Bregman method [19] to minimize u_1 and u_2. The global means c_i and variances σ_i^2 can be easily solved. For fixed $(u_1, u_2, c_i, \sigma_i)$, the bias field b can be determined by $b = \sum_{i=1}^{3}((\log I - c_i)M_i/\sigma_i^2)/\sum_{i=1}^{3}(M_i/\sigma_i^2)$. In view of the slowly varying property of the bias field, using the technique in [14], we can derive a smooth bias field in this form, $b = \Big(\big(\sum_{i=1}^{3}((\log I - c_i)M_i/\sigma_i^2)\big) * g\Big) \Big/ \Big(\big(\sum_{i=1}^{3}(M_i/\sigma_i^2)\big) * g\Big)$, where

$*$ is the convolution operation, and g is a lowpass filter, such as a mean-filter kernel or Gaussian-filter kernel.

After preliminary segmentation for CSF, we further apply the same scheme to segment the image into WM, GM and background by masking off the CSF. The energy is the same as Eq. (6). With these two convex models, we can achieve a good preliminary segmentation for the neonatal images. For example, Fig. 2(a) shows a slice of neonatal brain. Fig. 2(b) shows the preliminary segmentation result obtained by these convex models.

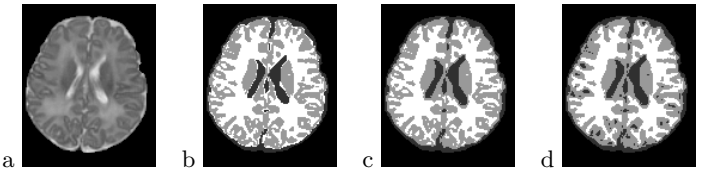


Fig. 2. Demonstration of preliminary tissue segmentation steps. (a) Original slice; (b) Preliminary segmentation; (c) PV removal; (d) Correction of CSF from WM.

2.3 Partial Volume (PV) Removal and CSF Correction from WM

After preliminary segmentation, we find that many voxels between CSF and GM are incorrectly classified as WM due to partial volume effect. In [2], Xue *et al.* proposed a technique based on EM algorithm and Markov random field (MRF) to remove the effect of PV. However, their PV removal strategy is somehow complicated. In this paper, we adopt rather simple but effective scheme to handle PV problem based on the observation that the misclassified WM are commonly surrounded by the CSF and GM. For each segmented WM voxel, in its neighborhood with size of $w \times w \times w$, let the number of WM, GM and CSF/BG be N_{WM}, N_{GM}, and N_{CSF}. If $N_{WM} \leq a$, while $N_{GM} > N_{CSF} \geq b$, then this WM should be set as GM. If $N_{WM} \leq a$, while $N_{CSF} > N_{GM} \geq c$, then this WM should be set as CSF. In this paper, we set $a = 3, b = 3, c = 6, w = 3$ for all experiments. Fig. 2(c) shows the correction result achieved by our PV removal scheme. In our experiment, we also find that there are some CSF voxels in sulci that are incorrectly labeled as WM. In this case, we adopt the scheme proposed in [2], which is based on the observation that these mislabeled CSF are unconnected with true WM volume, to correct these misclassified CSF from WM. Fig. 2(d) shows the correction result for CSF. This preliminary result will be used as a good initialization for the coupled level set method in Section 2.1.

3 Experimental Results

Data were acquired from a 3T Siemens scanner. T2 images of 70 axial slices were obtained with imaging parameters: TR=7380 ms, TE=119 ms, Flip Angle=150, acquisition matrix=256×128, and resolution=1.25×1.25×1.95 mm^3. T2 images

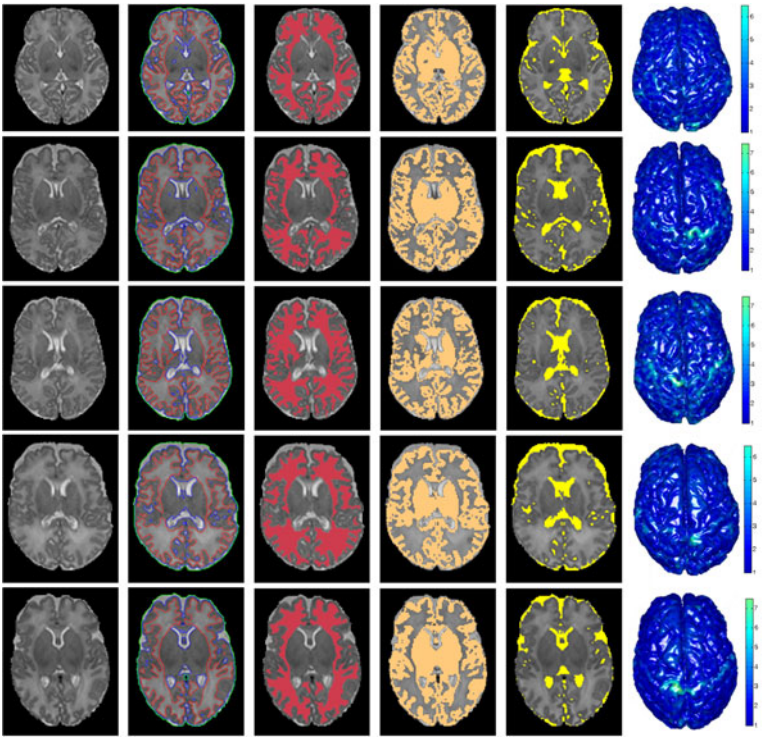
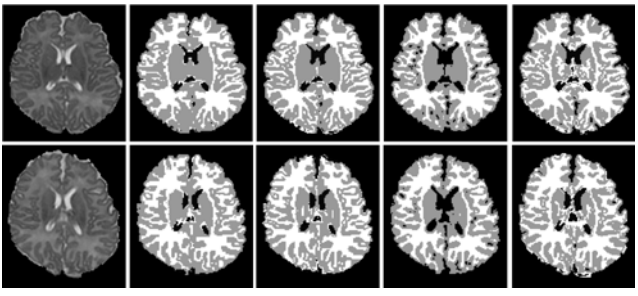


Fig. 3. 3D automatic neonatal brain segmentation results. From left to right: original T2 slices, results by the proposed method, segmented WM, GM, CSF, and thickness between WM/GM and GM/CSF surfaces.

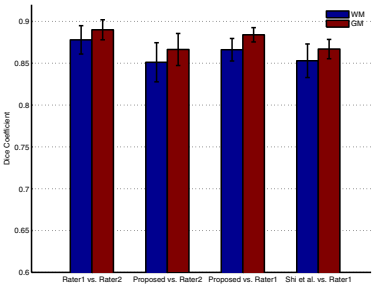
are resampled to $1\times1\times1$ mm^3 before further processing. In our experiments, we set the allowable thickness for cortex as [1 6.5]mm, $\nu = 0.5$ for Eq. (2), $\nu = 0.25$ for Eq. (6), and $\sigma = 3.0$. The functions δ and H are regularized as in [20]. The level set functions are reinitialized every iteration using fast marching method [21].

Our method has been validated using images obtained from 10 neonates. Due to the page limit, we show only the segmentation results for five subjects in Fig. 3. The first two columns show the original T2 slices and segmentation results of our coupled level set method. To better view the results, we also present the hard segmentations of WM, GM and CSF in columns 3, 4 and 5, respectively. Visual inspection of these results shows that WM, GM and CSF are reasonably well segmented. The last column visualizes the distance (or thickness) between WM/GM surface and GM/CSF surface. It can be seen that most of cortical thickness of these subjects are [1 5] mm, which is consistent with what we assumed.

Validation of the automatic segmentation results is difficult because ground truth is not available. For comparison, we have to refer to the manual



(a)



(b)

Fig. 4. (a): Segmentation results on two representative subjects. Column 1: original T2 images; columns 2 and 3: manual segmentation results by two experts; column 4: results by our proposed method; Column 5: results by the method of Shi *et al.*. (b): Comparison of segmentation accuracy on WM and GM for 10 subjects by different methods.

segmentations by experts as our ground truth. Fig. 4(a) shows the segmentation results of two representative subjects by two expert raters (columns 2 and 3) and by our method (column 4). By visual inspection, our results are comparable with those produced by expert raters. As atlas-based segmentation method is popular for neonatal brain segmentation, we do another comparison with the latest atlas-based segmentation method proposed by Shi *et al.* [3], with their results shown in the last column. To have a fair comparison for GM, we mainly compare the segmentation performance in the cortical regions. We employ Dice Coefficient (DC) [22] to measure the overlapping rate between two segmentations, which is defined as $DC = 2|A \cap B|/(|A| + |B|)$. DC ranges from 0 to 1, corresponding to the worst and the best agreement between labels of two regions. We first compare our results with manual segmentations by two raters. The mean and standard deviation of DC values of the WM and GM segmentations of all 10 subjects are presented in the first three pairs of bars in Fig. 4(b). We can observe that our proposed automatic segmentation method achieves comparable segmentation performance on WM and GM as manual raters. Taking the manual segmentation from rater 1 as ground truth, we then compare with the method of Shi *et al.* [3], and provide the results in the last pair of bars. It can be seen

that our proposed method outperforms the method of Shi *et al.*, by achieving the relatively higher DC values for both WM and GM.

4 Conclusion

We have presented a novel surface-based method for neonatal brain segmentation. Our method effectively utilizes local image information, atlas prior knowledge, and cortical thickness constraint for guiding the segmentation, by integrating them into a coupled level set method. We also provide a robust initialization method using convex optimization for this coupled level set method. Our proposed method has been validated on 10 subjects with promising results. In our future work, we will test our proposed method with more data.

References

1. Prastawa, M., Gilmore, J.H., Lin, W., Gerig, G.: Automatic segmentation of MR images of the developing newborn brain. Medical Image Analysis 9(5), 457–466 (2005)
2. Xue, H., et al.: Automatic segmentation and reconstruction of the cortex from neonatal MRI. NeuroImage 38(3), 461–477 (2007)
3. Shi, F., et al.: Neonatal brain image segmentation in longitudinal MRI studies. NeuroImage 49(1), 391–400 (2010)
4. Warfield, S.K., Kaus, M., Jolesz, F.A., Kikinis, R.: Adaptive, template moderated, spatially varying statistical classification. Medical Image Analysis 7(4), 43–55 (2000)
5. Weisenfeld, N.I., Warfield, S.K.: Automatic segmentation of newborn brain MRI. NeuroImage 47(2), 564–572 (2009)
6. Cocosco, C.A., Zijdenbos, A.P., Evans, A.C.: A fully automatic and robust brain MRI tissue classification method. Medical Image Analysis 7(4), 513–527 (2003)
7. Gooya, A., Liao, H., Matsumiya, K., Masamune, K., Masutani, Y., Dohi, T.: A variational method for geometric regularization of vascular segmentation in medical images. IEEE Transactions on Image Processing 17(8), 1295–1312 (2008)
8. Xu, C., et al.: Reconstruction of the human cerebral cortex from magnetic resonance images. IEEE Trans. Med. Imag. 18(6), 467–480 (1999)
9. Zeng, X., Staib, L., Schultz, R., Duncan, J.: Segmentation and measurement of the cortex from 3D MR images using coupled surfaces propagation. IEEE Trans. Med. Imag. 18(10), 100–111 (1999)
10. MacDonald, D., Kabani, N., Avis, D., Evans, A.C.: Automated 3-d extraction of inner and outer surfaces of cerebral cortex from MRI. NeuroImage 12(3), 340–356 (2000)
11. Goldenberg, R., Kimmel, R., Rivlin, E., Rudzsky, M.: Cortex segmentation: a fast variational geometric approach. IEEE Trans. Med. Imag. 21(2), 1544–1551 (2002)
12. Han, X., et al.: Cruise: Cortical reconstruction using implicit surface evolution. NeuroImage 23(3), 997–1012 (2004)
13. Wang, L., He, L., Mishra, A., Li, C.: Active contours driven by local gaussian distribution fitting energy. Signal Processing 89(12), 2435–2447 (2009)

14. Li, C., et al.: A variational level set approach to segmentation and bias correction of medical images with intensity inhomogeneity. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 1083–1091. Springer, Heidelberg (2008)
15. Li, C., Kao, C., Gore, J., Ding, Z.: Implicit active contours driven by local binary fitting energy. In: CVPR, pp. 1–7 (2007)
16. Paragios, N.: A variational approach for the segmentation of the left ventricle in mr cardiac images. In: VLSM 2001 (2001)
17. Bresson, X., et al.: Fast global minimization of the active contour/snake model. J. Math. Imaging Vis. 28(2), 151–167 (2007)
18. Chan, T.F., Esedoglu, S., Nikolov, M.: Algorithms for finding global minimizers of image segmentation and denoising models. SIAM J. Appl. Math. 66(5), 1632–1648 (2006)
19. Goldstein, T., Bresson, X., Osher, S.: Geometric applications of the split bregman method: Segmentation and surface reconstruction. CAM Report 09-06, UCLA (2009)
20. Chan, T., Vese, L.: Active contours without edges. IEEE Trans. Imag. Proc. 10(2), 266–277 (2001)
21. Sethian, J.: Level Set Methods and Fast Marching Methods. Cambridge University Press, Cambridge (1999)
22. Dice, L.: Measures of the amount of ecologic association between species. Ecology 26, 297–302 (1945)

A Unified Minimal Path Tracking and Topology Characterization Approach for Vascular Analysis

Hasan Ertan Çetingül[1], Mehmet Akif Gülsün[2], and Hüseyin Tek[2]

[1] Center for Imaging Science, Johns Hopkins University, Baltimore MD, USA
[2] Imaging and Visualization, Siemens Corporate Research, Princeton NJ, USA
ertan@cis.jhu.edu, {akif.gulsun,huseyin.tek}@siemens.com

Abstract. We present a unified coarse-to-fine approach for extracting the medial axis representations (centerlines) of human vasculature in contrast enhanced (CE)-CTA/MRA. The proposed method constitutes two separate analysis stages that are successively applied (and repeated) for a refined extraction. The former stage involves the use of a graph-based optimization algorithm that identifies the minimum-cost paths between user-specified seed points. The costs of all feasible paths are efficiently computed via the medialness filter, which is a contrast- and scale-invariant local operator sensitive to the presence of tubular structures. Nonetheless, image noise and the presence of nearby blood vessels can affect the quality of detection and delineation. In the latter stage, we thereby employ a novel multiscale orientation descriptor so as to guide/stop additional minimal path extraction steps. Specifically, the descriptor is designed to classify a point of interest as *vessel* or *non-vessel*, as well as to obtain a reliable estimate of the number and directions of the vascular segments (branches) at a vessel point. Our method improves the accuracy of extraction by robustly identifying critical configurations such as bifurcations, endpoints, or non-vessel points, and thereby delineating/eliminating missing/spurious vessel branches.

1 Introduction

In clinical applications, medial axis (centerline) representations of cardiac vasculature are important for diagnosing pathologies through advanced visualizations, treatment and surgery (e.g., CABG) planning, and follow-up studies. There has been numerous works in obtaining such representations from contrast enhanced (CE)-CTA/MRA (see [1] and references therein). In general, centerline segments between seed points can easily be extracted if the vessel segment is well-isolated or the user can place additional seed points at anatomically correct locations. However, in the case of more complex configurations, accurate, automatic and timely extraction of full centerline trees still remains a challenging problem.

A particularly popular class of algorithms known as the *minimal path* techniques identify the minimum-cost path as the vessel centerline. The costs of all feasible paths can be efficiently computed by using different operators such as the vesselness filters [2], medialness filters [3,4], image Hessian [5], polar intensity

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 11–20, 2010.

profiles [6], bank of oriented difference filters [7], or optimally oriented flux [8]. However, a resulting path cannot always correspond to the true centerline as there may be another path with a lower cost due to the presence of nearby structures. Furthermore, traditional minimal path schemes often face difficulties in detecting every vessel bifurcation when the goal is to extract the full centerline tree, e.g., coronary centerline tree. It is also difficult to determine the endpoints of the centerlines without producing additional spurious branches, i.e., poor convergence due to leakage towards the non-vascular areas [4].

In this paper, we thereby present a unified coarse-to-fine approach to extract the centerline representation of vessel trees robustly, timely and automatically. The proposed method involves the use of a minimal path-type tracking (propagation) algorithm, which employs an efficient contrast- and scale-invariant operator called the medialness filter for centerline localization. This stage can be considered as a traditional minimal path extraction scheme (described in our previous work [4]) to obtain a coarse estimate of the vessel centerline. The main contribution of this paper is to further improve the accuracy and robustness of extraction by characterizing image points on the propagating front via a multiscale orientation descriptor so as to guide/stop the propagation. Specifically, the descriptor is designed to classify a point of interest as *vessel* or *non-vessel*, as well as to yield a reliable estimate of the number and directions of the branches at a vessel point. Inspired by the concept of orientation distribution function (ODF) in diffusion weighted (DW)-MRI, the descriptor is used to compute a discrete spherical *configuration function* which provides, for each unit vector (point on the 2-sphere), an estimate of the probability of having a vascular structure oriented along that vector. The number of the modes of this function characterizes the configuration at the point of interest, whereas the modes themselves are used to guide the propagating fronts. We observe that successive application of this unified method (until the propagation stops) improves the quality of the extracted centerlines by identifying critical configurations, e.g., non-vessel points, bifurcations, vessel endpoints, and thereby extracting/eliminating missing/spurious branches.

2 Minimal Paths as Vessel Centerlines

Medical image analysis tools developed for clinical applications often employ graph-based minimal path extraction techniques for delineating centerlines of blood vessels [9,6,10]. Due to the existence of efficient implementation schemes, this type of algorithms provides an attractive approach to select the minimum-cost path among all feasible paths between user/automatically-placed seed points. A key ingredient in minimal path techniques is the definition (and use) of a local operator that computes the individual costs of moving from a voxel to another. The former stage of our proposed framework comprises such an algorithm, which employs multiscale medialness filters as the local operator.

Medialness filter: Vascular structures in CE-CTA/MRA have, in general, circular/elliptic shapes in cross-sectional views, e.g., a bright disk surrounded by a darker ring, even though the presence of nearby vessels or pathologies might

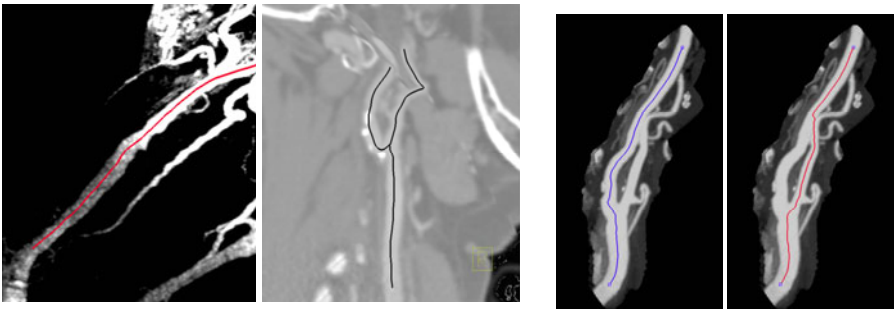
cause slight deviations in shape. The medialness filter is designed to capture this particular structure by means of computing several "edgeness" responses. Specifically, the medialness response $m(\boldsymbol{x})$ at a point $\boldsymbol{x}$ is computed on a circle of radius r centered at $\boldsymbol{x}$, i.e., $m(\boldsymbol{x}) = \max_{r\in\mathcal{R}} \sum_{n=0}^{N-1} \varepsilon(\boldsymbol{x} + r\boldsymbol{u}(2\pi n/N))$, where N defines the angular resolution, $\boldsymbol{u}(\alpha) = \sin(\alpha)\boldsymbol{e}_1 + \cos(\alpha)\boldsymbol{e}_2$ with $\boldsymbol{e}_1 \perp \boldsymbol{e}_2 \in \mathbb{R}^2$, $\|\boldsymbol{e}_k\| = 1$, for $k = 1, 2$, and ε measures the normalized edge response derived from the image gradients, as described in [4]. While the filter produces a strong response at the center of a vessel, the response drastically drops towards vessel boundaries and becomes very small (but non-zero) in non-vascular areas. However, one should also note that the value of the radius r is critical for decreasing the corruptive effect of nearby vessels or other bright structures.

Identifying paths from medialness maps: Having defined an operator to compute the medialness maps, we aim at finding the optimal feasible path (curve) between seed points, which is expected to coincide with the centerline of the vessel of interest. For this purpose, let $\mathcal{C}$ denote the set of all feasible curves between two seed points and $E(C)$ represent the *total energy* (or *cost*) along a curve $C(s)$ parameterized by the arc length s. The optimal curve $\hat{C}$ is identified as the curve with total minimum energy, i.e.,

$$\hat{C} = \arg\inf_{C\in\mathcal{C}} E(C) = \arg\inf_{C\in\mathcal{C}} \int_{\Omega} (P(C(s)) + \omega)\mathrm{d}s, \tag{1}$$

where $P(\cdot)$ is the *potential*, which corresponds to the inverse of the medialness measure at $\boldsymbol{x}$, i.e., $P(\boldsymbol{x}) = 1/m(\boldsymbol{x})$, and ω is the regularization term. Often encountered in various applications in computer vision, e.g., segmentation, this type of optimization problems can be solved by either Dijkstra's algorithm [11] or fast marching methods [12]. In this work, we employ Dijkstra's algorithm to solve (1) via explicit discrete front propagation. We use a 26-connected 3-D lattice and compute the cost of propagation between neighboring nodes from their potentials. The optimal curve is obtained by traversing (backtracking) along the propagation. We kindly refer the reader to [4] for an extended discussion and further details on the implementation.

The algorithm provides promising results even on vessels with high curvature or in the presence of strong calcification (see Fig. 1(a)). However, it should also be noted that it may not always yield the correct vessel centerline. For instance, as shown in Fig. 1(b), the algorithm might yield an erroneous path with a smaller cost. Since the medialness responses are low, yet non-zero around and outside the vessel walls, leakages towards the nearby vascular/non-vascular areas might occur. Furthermore, convergence of the propagating fronts, i.e., detection of vessel endpoints, might be problematic. In order to address these issues, one needs to obtain a more accurate, reliable and refined representation of the local vascular topology and the propagating fronts are remarkably well-positioned points of interest at which such local representations can be helpful. We thereby present the design of a multiscale orientation descriptor to obtain such a representation, i.e., configuration function, for eliminating spurious branches, guiding the propagating fronts towards vascular segments and determining their convergence.



(a) (b)

Fig. 1. Selected results of the minimal path detection algorithm: (a) Accurate centerline extractions, (b) Two distinct paths (red: *erroneous*, blue: *correct*) on a carotid artery in CE-CTA. The extraction is affected by the presence of a nearby structure.

3 Characterization of Local Vessel Configurations

In order to accurately guide the fronts towards vascular segments (or stop them generating spurious branches), we propose the use of a refined local representation, which we named as the *configuration function* (CF) $p : S^2 \mapsto \mathbb{R}^+$ computed via a robust multiscale orientation descriptor. The descriptor takes a spherical neighborhood around a point of interest $\boldsymbol{x}$ and estimates the value of $p(\boldsymbol{s}; \boldsymbol{x})$, i.e., the "probability" of having a linear vessel segment oriented along $\boldsymbol{s}$. For computational efficiency, we compute a discrete approximation of the CF at $N = 162$ vectors $\{\boldsymbol{s}_n\}_{n=1}^{N} \doteq \mathcal{S}$ defined by the twofold icosahedral tessellation.

3.1 Orientation Descriptor

As depicted in Fig. 2(a), the descriptor is an oriented matched filter centered at a fixed point $\boldsymbol{x}$ and has a moving point $\boldsymbol{f}$ located at a distance l from $\boldsymbol{x}$. Being a modified version of the filter presented in [13], it scans a candidate direction $\boldsymbol{s} \in \mathcal{S}$ by aligning the segment $\overline{\boldsymbol{xf}}$ with $\boldsymbol{s}$ such that $\boldsymbol{f} = \boldsymbol{x} + l\boldsymbol{s}$. The points $\{\boldsymbol{f}_k\}_{k=1}^{2K}$ are placed on a circle of radius r orthogonal to $\overline{\boldsymbol{xf}}$ so as to encapsulate the vessel segments around $\boldsymbol{x}$. Consecutive segments $\overrightarrow{\boldsymbol{xf}}_k$ and $\overrightarrow{\boldsymbol{xf}}_{k+1}$ are separated by an angle $\alpha = \pi/9$ and $(\boldsymbol{f}_k, \boldsymbol{f}_{k+K})$ forms an antipodal pair. Notice that the shape of the spatial support resembles a cylindrical tube oriented along $\boldsymbol{s}$ with length l and diameter $2r$, which is in accordance with the assumption of the aforementioned medialness filter on vessel cross-sections.

3.2 Computation of the Configuration Function

The descriptor is used to compute, $\forall \boldsymbol{s} \in \mathcal{S}$, three separate yet intertwined measures: *intensity homogeneity*, *minimum gradient flux*, and *nonlinear photometry*. Their combination gives an estimate of the CF at a point of interest.

Intensity homogeneity: The intensity values along a tubular structure are expected to be coherent. The first measure, similar to the one in [14], quantifies this homogeneity along $\boldsymbol{s} \sim \overline{\boldsymbol{xf}}$ as

$$H(\boldsymbol{s}, l; \boldsymbol{x}) \doteq \frac{1}{l} \int_0^1 |I(\boldsymbol{x} + \lambda l \boldsymbol{s}) - I(\boldsymbol{x})|^2 \mathrm{d}\lambda \approx \frac{1}{|\Omega_{\boldsymbol{s}}|} \sum_{\boldsymbol{p} \in \Omega_{\boldsymbol{s}}} |I(\boldsymbol{p}) - I(\boldsymbol{x})|^2, \qquad (2)$$

where $\Omega_{\boldsymbol{s}}$ denotes the set of voxels on $\overline{\boldsymbol{xf}}$. The homogeneity is maximized, i.e., $H(\boldsymbol{s}, l; \boldsymbol{x})$ is minimized, when $\boldsymbol{s}$ aligns with the vessel direction. The resulting measure is computed as $p_1(\boldsymbol{s}; \boldsymbol{x}) = \exp(-\beta_1 \times \sum_{l \in \mathcal{L}} H(\boldsymbol{s}, l; \boldsymbol{x}))$, where $\mathcal{L}$ is the set of feasible lengths and $\beta_1 > 0$ is a user-specified parameter.

Minimum gradient flux: This component uses a flux-type measure that seeks to align the normal vectors at $\{\boldsymbol{f}_k\}$ with the image gradients [15] as

$$F(\boldsymbol{s}, l, r; \boldsymbol{x}) \doteq \frac{1}{K} \sum_{k=1}^{K} \min\{|\boldsymbol{n}_k^\top \nabla I(\boldsymbol{f}_k)|, |\boldsymbol{n}_{k+K}^\top \nabla I(\boldsymbol{f}_{k+K})|\}, \qquad (3)$$

where $\boldsymbol{n}_k$ is the unit vector such that $\boldsymbol{f}_k = \boldsymbol{f} + r\boldsymbol{n}_k$, $\boldsymbol{n}_k = -\boldsymbol{n}_{k+K}$, and $\nabla I(\boldsymbol{p})$ is the image gradient at $\boldsymbol{p}$. Assuming that the vessels have circular cross-sections, F is maximized when $\{\boldsymbol{f}_k\}$ are placed on the vessel boundary. A coarse estimate of the radius of a vessel segment is thereby obtained as $\hat{r} = \mathrm{argmax}_{r \in \mathcal{R}} \sum_{l \in \mathcal{L}} F(\boldsymbol{s}, l, r; \boldsymbol{x})$. The resulting measure is computed as $p_2(\boldsymbol{s}; \boldsymbol{x}) = \exp(\beta_2 \times \sum_{l \in \mathcal{L}} F(\boldsymbol{s}, l, \hat{r}; \boldsymbol{x}))$ with a user-specified parameter $\beta_2 > 0$.

Nonlinear photometry: Here, the estimate of the radius is first used for repositioning the points $\{\boldsymbol{f}_k\}$ such that $r = \hat{r} + 1$, in order to *fully encapsulate* the vessel. These points are then utilized, together with $\{\boldsymbol{x}, \boldsymbol{f}\}$, to obtain the nonlinear photometry of the descriptor at $\boldsymbol{x}$. Specifically, for the k-th pair of antipodes $(\boldsymbol{f}_k, \boldsymbol{f}_{k+K})$, the partial photometry is computed as

$$D_k(\boldsymbol{s}, l; \boldsymbol{x}) = \begin{cases} 1 & \text{if } |I(\boldsymbol{f}) - I(\boldsymbol{x})| \leq \min\{|I(\boldsymbol{f}) - I(\boldsymbol{f}_k)|, |I(\boldsymbol{f}) - I(\boldsymbol{f}_{k+K})|\} \\ 0 & \text{otherwise.} \end{cases} \qquad (4)$$

The cumulative photometry is subsequently computed over all pairs of antipodes as $D(\boldsymbol{s}, l; \boldsymbol{x}) = \frac{1}{K} \sum_{k=1}^{K} D_k(\boldsymbol{s}, l; \boldsymbol{x})$. Notice that this response should be high when $\boldsymbol{s}$ aligns with the vessel direction. The resulting measure is computed as $p_3(\boldsymbol{s}; \boldsymbol{x}) = \exp(\beta_3 \times \sum_{l \in \mathcal{L}} D(\boldsymbol{s}, l; \boldsymbol{x}))$ with a user-specified parameter $\beta_3 > 0$.

The CF at a point $\boldsymbol{x}$ is computed by multiplying the aforementioned measures, i.e., $p(\boldsymbol{s}; \boldsymbol{x}) \propto \prod_{i=1}^{3} p_i(\boldsymbol{s}; \boldsymbol{x})$. By using such a combination strategy, one does not observe the sensitivity of the intensity homogeneity to noise or numerical errors in the computation of the image gradients for very thin vessels. However, each individual measure should have mode(s) (or relatively high values) at directions close to the ones of actual vascular segments in order not to cancel the contributions of the remaining measures out. Fig. 2(b) shows a bifurcating synthetic tubular structure with three points of interest and the resulting CFs at those points. Notice that the directions at which the CFs attain their modes coincide with (or are close to) the actual local orientations.

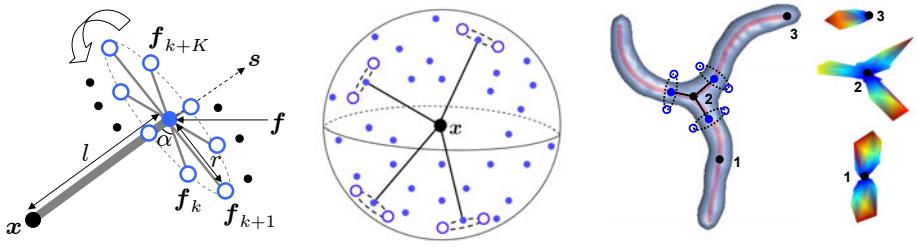


Fig. 2. (a) Support of the descriptor and its usage along candidate vectors, (b) 3 points of interest on a 3-D synthetic fiber (top view), the descriptor aligned with the vessel segments at point 2, and the CFs whose values are color-coded (blue∼low, red∼high)

3.3 Identification of Vessel Configurations

Characterization of the local vessel topology is performed by analyzing the CFs. If $\boldsymbol{x}$ is on a vessel, its CF $p(\cdot;\boldsymbol{x})$ is expected to have a "modal shape", otherwise the values $p(\boldsymbol{s}_n;\boldsymbol{x}) \approx 0, \forall n$. In addition, since $\{\boldsymbol{s}_n\} \subset S^2$ can be considered as data points with weights $\{p(\boldsymbol{s}_n;\boldsymbol{x})\}$, estimating the front propagation directions is equivalent to finding the modes of the CF. Therefore, one can apply the *uniformity test* [16] on spherical distributions to initially classify a point $\boldsymbol{x}$ as *vessel* or *non-vessel*, and then locate the modes of the CF if $\boldsymbol{x}$ is a vessel point.

Classification via the uniformity test: This test is primarily useful in identifying and pruning spurious branches. Let us denote the value $p(\boldsymbol{s}_n;\cdot)$ with ρ_n and write $\boldsymbol{s}_n = (x_n, y_n, z_n)$. The uniformity test involves the computation of the *resultant vector* $\boldsymbol{r} \doteq (\sum_n \rho_n x_n, \sum_n \rho_n y_n, \sum_n \rho_n z_n)$. The reliability of the test can be further improved by defining a *brightness factor* γ to modify the resultant vector such that $\boldsymbol{r} \leftarrow \gamma \boldsymbol{r}$. Here, the factor γ emphasizes locally brighter voxels, which is a common occurrence for blood vessels in CE images. It is computed as

$$\gamma \doteq [1+\exp(-\beta_4 \times |\max\{I(\boldsymbol{p}) : \boldsymbol{p} \in \mathcal{P}_{\boldsymbol{s}_{H_{\min}}}\} - \min\{I(\boldsymbol{p}) : \boldsymbol{p} \in \mathcal{P}_{\boldsymbol{s}_{H_{\max}}}\}|)]^{-1}, \quad (5)$$

where $\beta_4 > 0$ is a user-specified parameter, $\mathcal{P}_{\boldsymbol{s}_{H_{\min}}}$ and $\mathcal{P}_{\boldsymbol{s}_{H_{\max}}}$ are the sets of voxels along the direction of minimum and maximum homogeneity, respectively. The uniformity test proceeds as follows: If $\|\boldsymbol{r}\| \approx 0$, then the vectors $\{\boldsymbol{s}_n\}$ with weights $\{\rho_n\}$ are "uniformly" distributed on S^2 and the point of interest is identified as *non-vessel*. However, if $\|\boldsymbol{r}\| \gg 0$, the corresponding CF has mode(s) and the point of interest is on a vascular segment. For the latter case, we further identify the type of vascular topology by detecting the mode(s) of the CF.

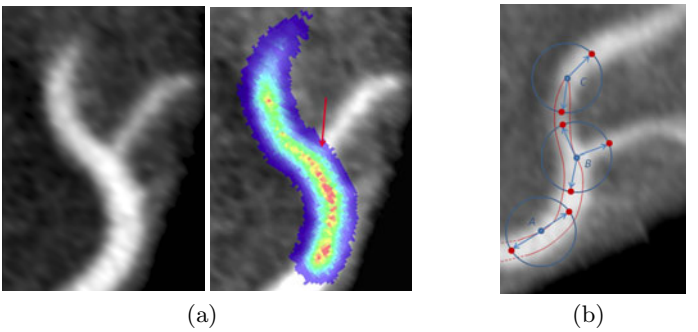
Mode detection via spherical clustering: The mode detection problem is solved by employing a nonparametric kernel density estimator called the *mean shift* algorithm [17]. The kernel is selected to be the von Mises-Fisher kernel $\Phi(\boldsymbol{s}, \boldsymbol{\mu}; \kappa)$ [16] between $\boldsymbol{\mu}$ and $\boldsymbol{s}$ with the concentration parameter $\kappa > 0$. The method approximates a CF at $\boldsymbol{\mu}$ as $\tilde{p}(\boldsymbol{\mu};\boldsymbol{x}) = Z_\Phi \sum_n \frac{\rho_n \kappa_n}{4\pi \sinh(\kappa_n)} \exp(\kappa_n \boldsymbol{s}_n^\top \boldsymbol{\mu})$

with a normalization term Z_Φ [13]. The adaptive factor κ_n is computed as the inverse of the geodesic distance between $\boldsymbol{s}_n$ and its third nearest neighbor and the modes are located by gradient ascent. As a result, the algorithm can, for instance, further identify a vessel point as a *bifurcation* (trimodal CF), a *regular point* (bimodal CF), or an *endpoint* (unimodal CF) (see Fig. 2(b)).

4 Unified Approach: Overview and a Motivating Scenario

The propagation algorithm employing the medialness filter produces promising trajectories, which often coincide with the actual vessel centerlines. Nonetheless, longer propagations to locate the branch endpoints may yield anatomically incorrect results as well as unnecessarily increased computation times due to the accumulative nature of the method. Moreover, in the case of bifurcations, the fronts should propagate towards the vessel walls to track both branches. However, the filter response might be low when a branch is very thin relative to the other (Fig. 3(a)). We address these issues by integrating the aforementioned topology characterization stage into the propagation scheme.

Specifically, we apply the orientation descriptor at the front points, denoted by $\mathcal{X}$, to either guide the propagation solely towards unexplored vessels/branches or to terminate it at a branch endpoint. In other words, *if a front point* $\boldsymbol{x} \in \mathcal{X}$ *is identified as non-vessel, we remove* $\boldsymbol{x}$ *from* $\mathcal{X}$ *and terminate its propagation.* Alternatively, *if* $\boldsymbol{x}$ *is identified as a vessel point, we find the direction(s) at which the CF* $p(\cdot;\boldsymbol{x})$ *attains its mode(s) and favor the propagation towards the "unexplored" direction(s).* We demonstrate this procedure in Fig. 3(b), which shows three points of interest $\{A, B, C\}$ on the moving front of an artery. Accordingly, the front point A is removed from the propagation while B and C are kept for further propagation as the CFs at these points indicate the presence of possibly unexplored areas. Likewise, the orientation descriptor at the endpoint of a vessel branch would yield a unimodal CF whose mode indicates an already-explored area, generating a termination flag for the algorithm.



(a) (b)

Fig. 3. (a) Response of the medialness filter (blue∼low, red∼high) along a bifurcating vessel, (b) Local orientations extracted via the descriptor at points $\{A, B, C\}$

5 Validation and Discussions

Validation of the proposed method is performed via two different experiments. We initially evaluate the stand-alone performance of the orientation descriptor at selected points of interest in CTA data. Recall that the support of the descriptor should be constructed to fully encapsulate the coronary arteries. Therefore, after a careful examination of the structure (width and curvature) of the arteries in the dataset, the support is formed[1] such that $l \in \{0.3, 0.6, \ldots, 3\}$ and $r \in \{0.4, 0.8, ..., 2\}$ (in voxels). It is also worth noting that the maximum value of l determines the extent of "multiscale" characteristic of the descriptor. Finally, the values for β_i, which describe the amount of deviation from the modes, are set to $\{\beta_i\}_{i=1}^{4} = \{6, 3, 3, 10\}$. Although the performance of the descriptor does not change drastically depending on the values of β_i, a good rule of thumb would be to select values between 1 to 20. Fig. 4 shows the outputs of the descriptor at several points of interests (in blue) placed on/outside of two coronary arteries. While the trajectories in yellow delineate the actual centerlines, the points in red are placed to indicate the local vessel directions, i.e., vectors from blue points to nearby red points. We obtained such promising results that for the selected arteries, the descriptor achieves an excellent performance by misidentifying only one bifurcation. Overall, it correctly classified about 90% of the analysis points in the dataset and different parameter settings produced comparable results.

We subsequently tested the proposed unified approach on several coronary CTA data where the ostia points are automatically detected for algorithm initialization. Fig. 5 shows two automatically constructed coronary centerline trees using this approach. We observed that the method successfully detected clinically significant branches and decreased the number of spurious branches. In addition, small distal centerlines were detected due to improved convergence. Another advantage of the algorithm is that since a vessel bifurcation can be detected at one of the many moving front points, missing the bifurcation at a single front point does not affect the extraction accuracy. Due to strong empirical evidence, we anticipate that the proposed approach quantitatively outperforms its predecessor in [4], whose performance was reported in great detail in [1].

In conclusion, the integration of the estimation and analysis of the configuration function into the minimal path detection algorithm improves the accuracy of vessel centerline tree extraction by identifying critical configurations such as bifurcations, endpoints, and eliminating erroneous vessel branches. Specifically, the unified approach provides the following improvements: 1) leakage to neighboring vessels or other bright structures are often prevented, 2) automatic convergence of the propagating fronts are established, and 3) branching topologies are identified more accurately. Our future work will focus on achieving such results in real-time. In addition, other supervised mode detection strategies such as k-means or expectation-maximization will be tested as the local topology of the arteries is anatomically restricted.

[1] An alternative strategy to adjust such parameters is to perform statistical inference from training data.

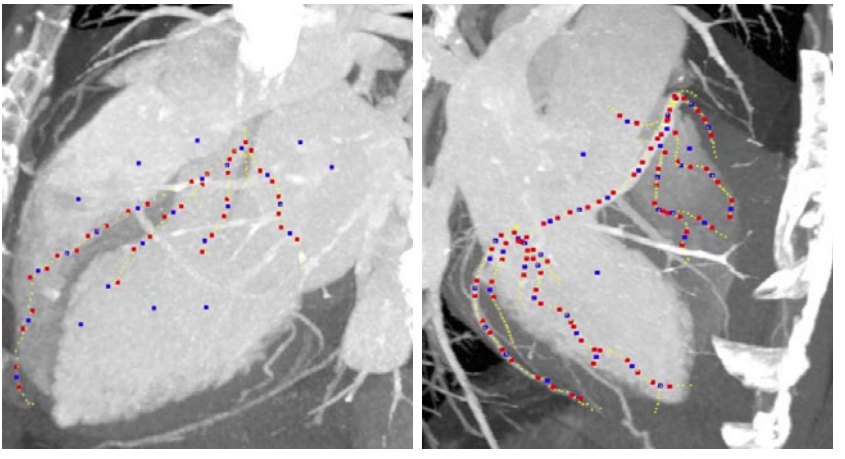


Fig. 4. Evaluation of the orientation descriptor on selected arteries: Ground truth centerline (yellow), analysis points (blue), resulting vessel/branch directions (blue→red)

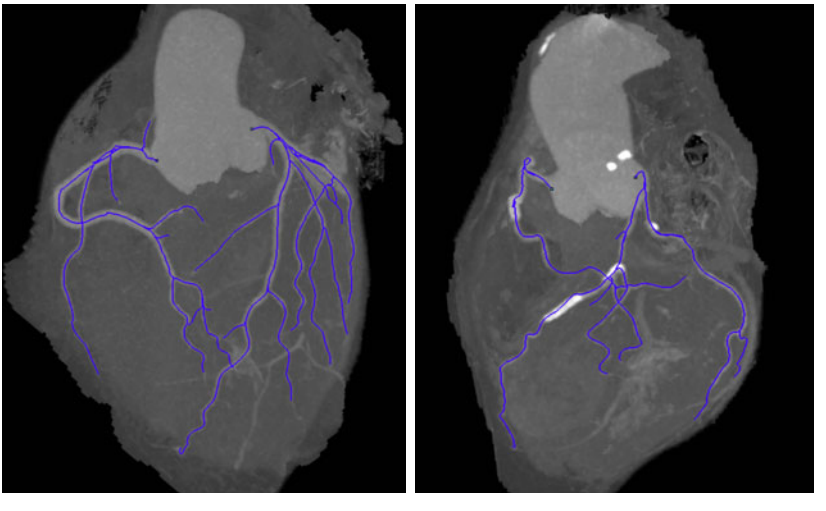


Fig. 5. Extracted coronary centerline trees using the proposed unified approach (initiated at automatically detected ostia points)

References

1. Schaap, M., et al.: Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms. Medical Image Analysis 13(5), 701–714 (2009)
2. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale vessel enhancement filtering. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 130–137. Springer, Heidelberg (1998)

3. Alyward, S., Bullitt, E.: Initialization, noise, singularities, and scale in height ridge traversal for tubular object centerline extraction. IEEE Trans. on Medical Imaging 21, 61–75 (2002)
4. Gülsün, M.A., Tek, H.: Robust vessel tree modeling. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 602–611. Springer, Heidelberg (2008)
5. Wink, O., Niessen, W., Viergever, M.: Multiscale vessel tracking. IEEE Trans. on Medical Imaging 23(1), 130–133 (2004)
6. Li, H., Yezzi, A.: Vessels as 4-D curves: Global minimal 4-D paths to extract 3-D tubular surfaces and centerlines. IEEE Trans. on Medical Imaging 26(9), 1213–1223 (2007)
7. Péchaud, M., Keriven, R., Peyré, G.: Extraction of tubular structures over an orientation domain. In: IEEE Int. Conf. on Computer Vision and Pattern Recognition, pp. 336–342 (2009)
8. Benmansour, F., Cohen, L., Law, M., Chung, A.: Tubular anisotrophy for 2D vessels segmentation. In: IEEE Int. Conf. on Computer Vision and Pattern Recognition, pp. 2286–2293 (2009)
9. Deschamps, T., Cohen, L.: Fast extraction of minimal paths in 3D images and applications to virtual endoscopy. Medical Image Analysis 5(4), 281–299 (2001)
10. Tyrrell, J., et al.: Robust 3-D modeling of vasculature imagery using superellipsoids. IEEE Trans. on Medical Imaging 26(2), 223–237 (2006)
11. Dijkstra, E.: A note on two problems in connections with graphs. Numerische Mathematic 1, 269–271 (1959)
12. Sethian, J.A.: Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science, 2nd edn. Cambridge University Press, New York (1999)
13. Çetingül, H., Plank, G., Trayanova, N., Vidal, R.: Estimation of multimodal orientation distribution functions from cardiac MRI for tracking Purkinje fibers through branchings. In: IEEE Int. Sym. on Biomedical Imaging, pp. 839–842 (2009)
14. Qian, X., et al.: A non-parametric vessel detection method for complex vascular structures. Medical Image Analysis 13, 49–61 (2009)
15. Lesage, D., Angelini, E., Bloch, I., Funka-Lea, G.: Design and study of flux-based features for 3D vascular tracking. In: IEEE Int. Sym. on Biomedical Imaging, pp. 286–289 (2009)
16. Fisher, N., Lewis, T., Embleton, B.: Statistical analysis of spherical data. Cambridge University Press, Cambridge (1993)
17. Comaniciu, D., Meer, P.: Mean Shift: A robust approach toward feature space analysis. IEEE Trans. on Pattern Analysis and Machine Intelligence 24(5), 603–619 (2002)

Subject Specific Shape Modeling with Incremental Mixture Models

Lichao Wang[1], Karim Lekadir[1], Ismail EI-Hamamsy[2], Magdi Yacoub[2], and Guang-Zhong Yang[1]

[1] Institute of Biomedical Engineering, Imperial College London, UK
[2] National Heart and Lung Institute, Imperial College London, UK

Abstract. Statistical shape models provide versatile tools for incorporating statistical priors for image segmentation. Difficulties arise, however, when the target anatomical shape differs significantly from the training set used for model construction. This paper presents a novel approach for fast and accurate segmentation of subject-specific geometries based on models largely derived from normal subjects. This technique is particularly suitable for analyzing complex structures such as severely abnormal patient datasets. The proposed method uses online principal component update to incorporate subject-specific geometry. Mixture models are used to estimate the latent density distribution of the data, thus enabling adequate constraining during active shape propagation. Validation based on hypertrophic cardiomyopathy (HCM) datasets with MRI shows significant improvement in overall accuracy and increased adaptation to complex structures.

Keywords: Statistical shape models, subject-specific segmentation, incremental principal component analysis, Gaussian mixture models.

1 Introduction

In medical image computing, statistical shape modeling [1] has an established role for studying complex 3D geometries. One of the main challenges of the technique is in the handling of unseen shapes, particularly for morphological variations associated with pathological changes. For patients with Hypertrophic Cardiomyopathy (HCM) – a common form of genetic myocardial disease, for example, a large part of the myocardium may be thickened at the septal region, thus causing significant deviations to normal cardiac morphology. To compensate for the associated functional abnormality, the left ventricle undergoes significant shape remodeling over time at different parts of the myocardium. In particular, the endocardium can lose its sphericity and the remodeled morphology tends to be arbitrarily complex [2]. In such situations, the application of statistical shape modeling can result in poor approximation of what is often an unpredictable structure based on the captured modes of variation of the training data. To cater for better approximation of unseen shapes, a number of shape modeling techniques have been proposed to improve the capabilities of statistical models built from limited training samples. In [3], for example, the use of wavelet transforms

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 21–30, 2010.

as the underlying variables has been attempted to alleviate the model over-fitting problem in active shape models (ASMs). In [4], synthetic models were combined with statistical modes of variation to allow for additional flexibility during segmentation. An alternative solution was developed in [5], suggesting the enlargement of the training set by the incorporation of artificial variations. These techniques, however, cannot handle the case of arbitrarily abnormal shapes such as the ones encountered in HCM datasets, which tend to differ significantly from subject to subject, even among the same family. Consequently, a more practical approach is to incorporate adaptive (incremental) training during segmentation such that a certain amount of information about the subject-specific geometry can be learnt. In [6], for example, the authors propagate the end-diastolic boundaries across the cardiac cycle by using a multi-linear dynamic model. The technique, however, is not designed for analyzing severely abnormal cases.

In this paper, a method is introduced for efficient incorporation of subject specific information for accurate spatio-temporal segmentation. The technique is particularly suitable for complex and diseased anatomical structures that are described by cine image data. An incremental principal component algorithm is presented, which enables fast and efficient modification of the main modes of shape variation. The gathered knowledge is then incorporated into a mixture model for improving the constraints used for segmenting subsequent image data. In essence, this enables the new model to combine both the temporal properties as captured offline from the initial training data and subject-specific properties online. The framework is applied in this paper to MR segmentation of the endocardial boundaries of patients with HCM. Detailed experiments are carried out to quantify the performance achieved by the proposed incremental mixture model (IMM).

2 Methods

2.1 Incremental PCA

One of the key considerations of this paper is on efficient online learning of complex shapes. For normally distributed samples, Principal Component Analysis (PCA) is widely used for statistical shape modeling. This is not appropriate for online dimensionality reduction since this would involve the computation of a new set of eigenvectors from a growing training set. In this paper, an update scheme based on Incremental PCA (IPCA) is used instead. IPCA and Incremental Singular Value Decomposition (ISVD) [7] are effective methods for estimating missing or contaminated information by back projecting the data from the PCA space to the original axes. In this paper, a more robust IPCA method as recently proposed in [8] is used, which involves the following key steps. Firstly, in order to obtain the residuals r associated with a new shape x, its closest instance within the current model is computed as follows:

$$y = U_n a + \mu_n \tag{1}$$

where $a = U_n^T(x - \mu_n)$, $r = x - y$, and U_n is the current eigenvector matrix and μ_n the mean of all the batch samples. In Eq. (1), x is the new shape, y represents the closest shape vector within the model. Subsequently, the eigenvector matrix is increased by one dimension before rotation, *i.e.*,

$$U' = \begin{pmatrix} U_n & \frac{r}{\|r\|} \end{pmatrix} \tag{2}$$

where U' is the appended eigenvector matrix. The dimensionally-increased eigenvector matrix is in the same form as a real eigenvector matrix, except that it does not balance the new data coefficients. To this end, PCA is performed on the appended coefficient matrix with the aim to obtain the rotation matrix from the principal components, *i.e.*,

$$A' = \begin{pmatrix} A_n & a \\ 0 & \|r\| \end{pmatrix} \xrightarrow{\text{PCA}} \mu'', U'', \lambda'' \tag{3}$$

where A' and A_n are the appended and current coefficient matrices in the eigenspace, respectively. μ'', U'' and λ'' are the mean, eigenvectors and eigenvalues of A', respectively. The new eigenvectors of the balanced system are those in the dimension-increased eigenvector matrix after rotation, *i.e.*, $U_{n+1} = U' \cdot U''$ and $\mu_{n+1} = \mu_n + U'\mu''$, where U_{n+1} and μ_{n+1} are the rotated eigenvectors and the updated mean, respectively. The least significant principal components are then removed as in conventional PCA. In this study, user delineations are incrementally added into the new model by using the IPCA as described above. As a result, the space is balanced as the variation energy of the newly added shapes in the PCA space is reduced.

2.2 Incremental Mixture Models

In practice, the delineated samples can have a large distance to the mapped training data in the multi-dimensional feature space due to the very nature of subject-specific modeling. In this case, the PCA axes would attempt to interpolate this new non-Gaussian distribution, thus introducing a bias in the allowable domain that will prohibit suitable use of the incorporated subject-specific information. To circumvent this problem, we have combined IPCA with a mixture model approach [9]. With this method, the probability associated with a given shape is computed and if it is lower than a predefined threshold, the shape is regarded as invalid even if it is within the original allowable domain as defined by PCA. With this mixture model, the shape regulation can be regarded as a multi-class classification problem, where the membership of a shape belonging to a certain cluster of shapes can be computed by using Bayes rule. In this way, the density estimation based on training data is a maximum likelihood estimation (MLE) problem [10]:

$$\hat{f}(x;\phi) = \sum_{i=1}^{g} \pi_i f_i(x;\theta) \tag{4}$$

where $\hat{f}$ is the estimated density of the mixture models, f_i is the ith model, π_i is the weight of the ith model, θ is the parameters of the model, g is the total number of models, and ϕ are the nuisance parameters to be estimated. By choosing a Gaussian kernal for the density estimation of the multivariate data, Eq. (4) becomes:

$$\hat{f}(x) = \sum_{i=1}^{g} \pi_i N(x : \mu_i, S_i) \tag{5}$$

and the nuisance parameters are the weights, mean μ_i and variance S_i of all the Gaussian bumps. To compute the MLE, ideally marginal likelihood functions should be known. The estimation of these functions is performed through an expectation-maximization (EM) algorithm [9], *i.e.*,

$$S_i = \frac{1}{Nw_j} \sum p_{ij} \left[(x_i - \mu_j)(x_i - \mu_j)^T + T_i \right] \tag{6}$$

$$p_{ij} = \frac{w_j N(x_i : \mu_j, S_j)}{\sum_{j=1}^{m} w_i N(x_i : \mu_j, S_j)} \text{ and } w_j = \frac{1}{N} \sum_i p_{ij}, \mu_j = \frac{1}{Nw_j} \sum_i p_{ij} x_i \tag{7}$$

where T_i is a fixed covariance of each sample.

Based on the incremental principal components and the mixture models described above, the implementation of IMM is as follows: by incorporating the results from IPCA, the starting position of the modified EM algorithm becomes the updated eigen-coding rather than the batch eigen-coding. The means of Gaussian clusters are IPCA eigen-codings of randomly selected shapes in the training set, *i.e.*, $\mu_i = T_{IPCA}(s_j)$ and $S_i = T_j$, where i and j denote the ith Gaussian bump and the jth sample, respectively, and T_{IPCA} denotes the IPCA transformation as described in section 2.1 while s_j is the jth shape in the training set.

To perform segmentation, we first use all shapes in the training set to build a statistical shape representation based on the mixture models. The new subject-specific shapes are then incorporated by updating the original components using IPCA. Subsequently, the EM algorithm as described above is used to generate the density of latent variables, with the starting positions obtained from IPCA update. For a new segmentation task, the boundary points detected through appearance matching on the image data are projected into the updated PCA space. If the probability of the projected shape is above the predefined threshold, it is considered as valid and can be used for the next iteration of appearance matching and regulation. If the probability is below the threshold, the shape is classified to the closest shape cluster. It then moves uphill in terms of the cost function towards the relevant cluster until its probability is greater than the threshold and thus the regulated shape can be used for subsequent active search iteration. It is worth noting that an appropriate threshold can be chosen as a factor (generally chosen between 2 and 3) of the standard deviation of the lowest Gaussian bump.

The proposed incremental mixture model has a number of advantages for spatio-temporal tracking of abnormal morphology. Unlike approaches based on the ASM [1] or batch mixture models [9], it can efficiently update the statistical model with subject-specific information that can guide the spatio-temporal tracking towards more plausible solutions. Additionally, while the incremental PCA proposed in [11] uses a multivariate Gaussian hypothesis that promotes solutions closer to the normal training data, IMMs are particularly suitable for the segmentation of outlying test shapes (*e.g.*, severe abnormal data) as illustrated in Fig. 4, where it can be seen that they can constrain subsequent tracking accurately.

2.3 Data Collection

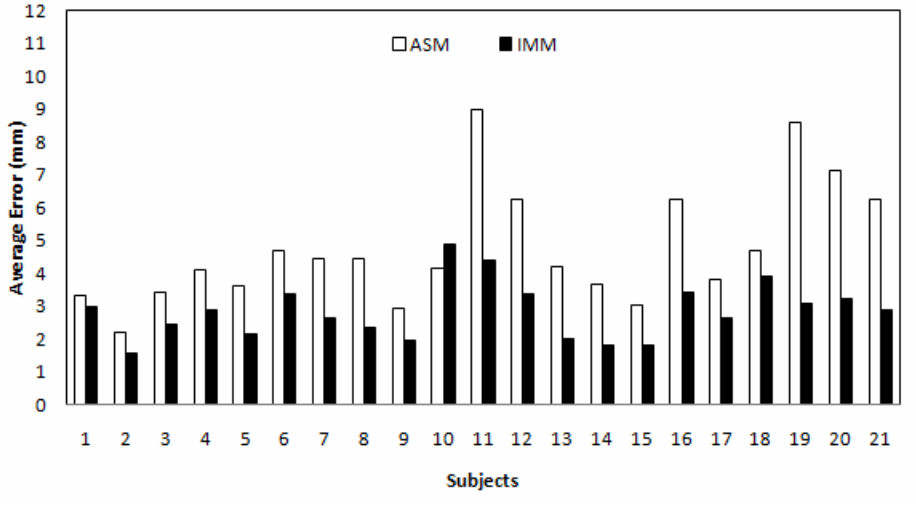
The left ventricular samples used for validation were collected by scanning a total of 81 subjects (60 normal and 21 patients with HCM) using a 1.5T MR scanner (Siemens Sonata 1.5T, Erlangen, Germany) and a trueFISP sequence. The acquisition parameters consist of an in-slice pixel resolution between 1.5 and 2 mm, a slice thickness of 10 mm, and a temporal resolution from 31.5 ms to 37.8 ms. For deriving the ground truth for the endocardial boundaries of the LV, manual delineation was performed by an expert clinician using 128 landmarks for each 3D shape. Additionally, the RV/LV junction points were manually defined in all short-axis images and frames. Based on these positions, the landmarks were uniformly distributed along the boundaries to establish point correspondences. Finally, the image datasets were temporarily re-sampled so that they have approximately a temporal resolution of 48 ms.

3 Results

The proposed framework was applied for the segmentation of the 21 HCM cine datasets, where the batch model was constructed based on the 60 normal datasets. For comparison, three different versions of the ASM that include the original formulation in [1], the Gaussian mixture model technique (MM) [9] and the IPCA in [11] were implemented. Both the ASM and the MM techniques utilized the batch model for the segmentation, while the IPCA and the IMM performed a model update at each search based on the shape at previous frame Therefore, the user was only required to delineate the first frame in order to initiate the segmentation process for the entire cardiac cycle. All the segmentation techniques were initialized at each frame by placing the 3D shape obtained at previous frame at the center of the target LV. The 4D tracking was carried out starting at end-diastole until the end-systolic time frame. For all four methods, 30 principal components were selected to form the feature space. For the modeling and evaluation of grey-level local appearance, standard eigen-profiles were used following the method by Cootes *et al.* in [12]. We have focused in this paper on the endocardial boundaries because they are geometrically challenging for HCM datasets due to the significant remodeling and complexity involved. Validation with other cardiac structures, *e.g.*, LV epicardial borders, is part of our future work.

Table 1. Segmentation error statistics for the ASM, MM, IPCA and IMM (in *mm*)

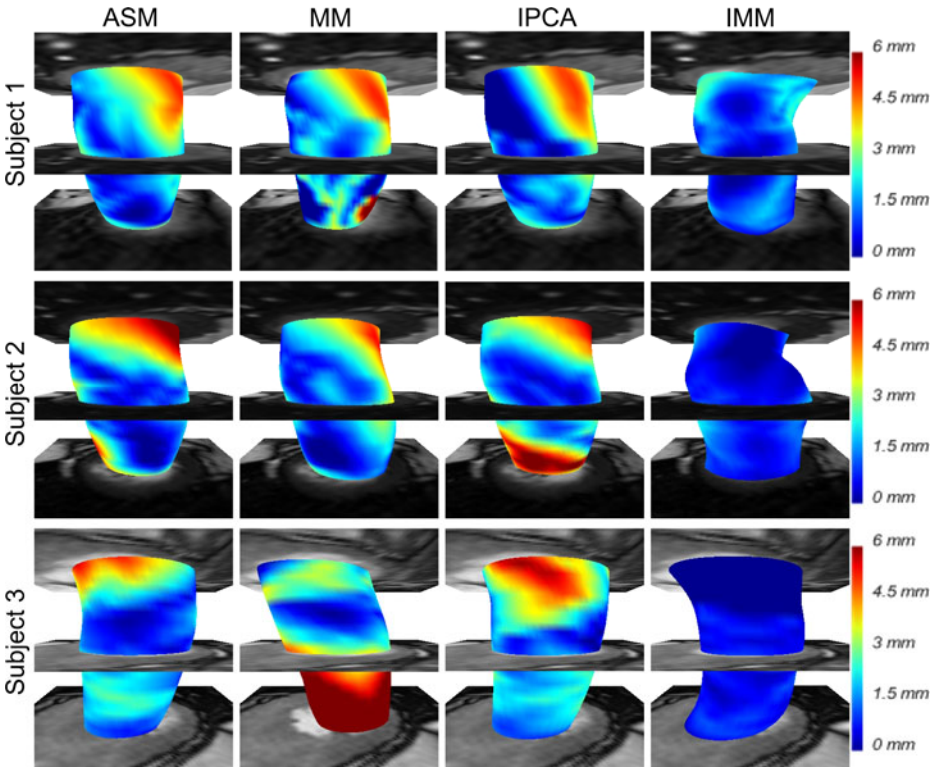
	ASM	MM	IPCA	IMM
Mean error	4.78	4.90	4.98	2.86
St. deviation	1.84	1.31	2.05	1.90



Flg. 1. Detailed segmentation results for the 21 HCM datasets as obtained by using the ASM and the proposed IMM framework

Table 1 summarizes the mean point-to-point errors and standard deviations for all the techniques used for comparison. For a more detailed evaluation of the results, the segmentation errors for the 21 HCM data are displayed in Fig. 1 for the IMM and the ASM. It can be seen from the results that the proposed framework introduces a marked improvement for most datasets, with an average improvement of 40% over the original ASM. This performance is particularly significant given the complex geometry and motion associated with HCM datasets. This is clearly evident in the examples of Figs. 2 and 3, where the target endocardial structures display severely abnormal morphology due the cardiac remodeling. As expected, the original ASM over-constraints the search procedure, thus generating new instances that distance themselves from the image data. The same problems can be seen with the MM and IPCA techniques, which are affected by the severely abnormal morphologies. The proposed technique, however, allows for more flexibility during shape localization through improved projection onto the new model components. These results demonstrate the clinical potential of the technique, since manual delineation of a single frame is sufficient to obtain reasonably accurate segmentations of datasets as challenging as those of HCM patients.

Fig. 4 illustrates the benefits of the proposed technique for the segmentation of severely abnormal data. It can be seen from the projection onto a shape subspace that

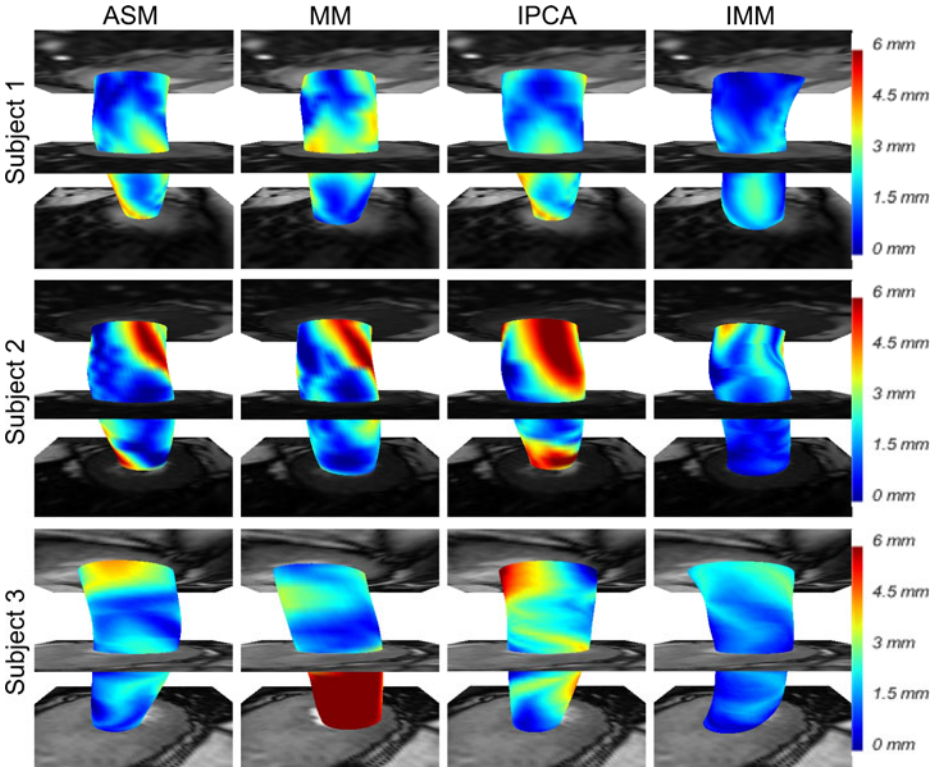


Flg. 2. Three illustrations at 142 ms of the cardiac cycle showing surface localization errors for the four methods used for comparison. With the proposed technique, marked improvement is achieved throughout the entire surfaces.

the target shape (blue triangle) is significantly different from the normal training samples (red circles). When updating the model based on the shape at previous frame (blue circle) and by using the IPCA technique in [11], the allowable shape space is still biased towards the training data (ellipse). As a result, the associated segmentation fails to recover the true boundaries (red star). With the proposed technique, however, the use of incremental mixture models means the segmentation is guided towards a more plausible solution and the output segmentation (blue star) is very close to the target HCM shape.

A number of points related to the implementation of the proposed technique need to be discussed. Firstly, the probability rejection threshold in the IPCA space is a key factor for the performance of the proposed algorithm. In practice, choosing a very high value can lead to over-fitting problems, while a small threshold can be vulnerable to noise. We found that setting the probability rejection threshold to the probability at two times the standard deviation of the lowest Gaussian bump in the mixture model leads to satisfactory results. A more systematic study of the sensitivity of this parameter is part of our future work. Moreover, the off-line standard EM algorithm used in this paper is inherently time consuming, thus prohibiting rapid segmentation

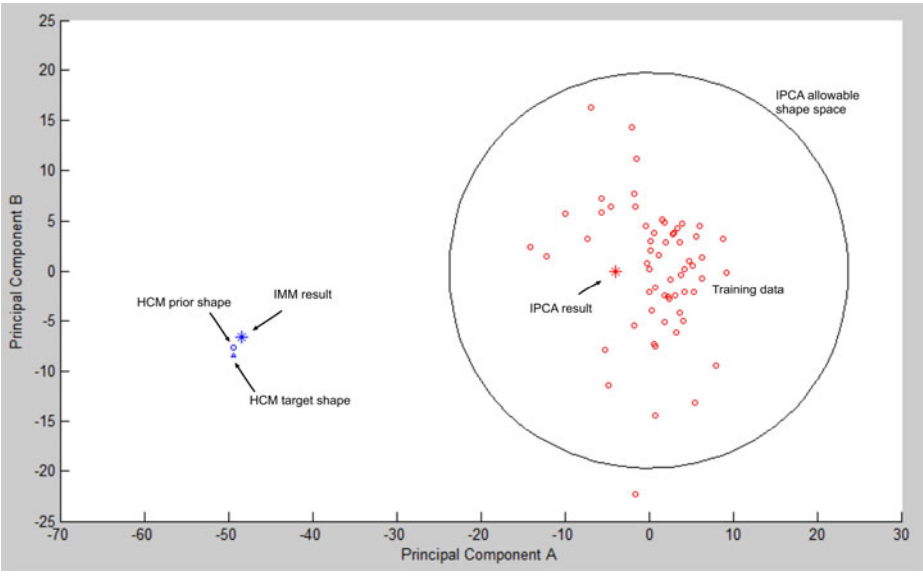
and data analysis. However, online versions of the EM algorithm have been recently developed (*e.g.*, [13]), which can be implemented as an alternative for faster segmentation using the proposed framework.



Flg. 3. Three illustrations at 332 ms of the cardiac cycle showing surface localization errors for the four methods used for comparison

4 Conclusion

In this paper, we have presented a robust segmentation scheme that provides efficient online update of the statistical shape model. The framework enables the incorporation of subject-specific information that is otherwise difficult to obtain through prior training sets. This is achieved by an incremental mixture model that permits the analysis of plausibility measures for new shapes based on the latent data. The constraints provide improved adaptation to complex shapes and permit more accurate tracking of the boundaries across the cardiac cycle. Validation with HCM left ventricular datasets demonstrates the potential clinical value of the technique.



Flg. 4. Illustration of the adaptability of the proposed IMM method to arbitrarily complex shapes. Unlike the IPCA which is biased by the normal subjects despite the model update, the proposed IMM is guided towards a solution in the shape space that is more plausible.

References

1. Cootes, T.F., Taylor, C.J.: Active shape models. In: British Machine Vision Conference, pp. 266–275 (1992)
2. Cecchi, F., Yacoub, M.H., Olivotto, I.: Hypertrophic cardiomyopathy in the community: why we should care. Nature Clinical Practice Cardiovascular Medicine 2, 324–325 (2005)
3. Davatzikos, C., Tao, X., Shen, D.: Hierarchical active shape models, using the wavelet transform. IEEE Transaction on Medical Imaging 22, 414–423 (2003)
4. Wang, Y., Staib, L.H.: Boundary finding with prior shape and smoothness models. IEEE Transaction on Pattern Analysis and Machine Intelligence 22, 738–743 (2000)
5. Lötjönen, J., Antila, K., Lamminmäki, E., Koikkalainen, J., Lilja, M., Cootes, T.F.: Artificial enlargement of a training set for statistical shape models: Application to cardiac images. In: Functional Imaging and Modelling of the Heart, FIMH (2005)
6. Zhu, Y., Papademetris, X., Sinusas, A.J., Duncan, J.S.: Bidirectional segmentation of three-dimensional cardiac MR images using a subject-specific dynamical model. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 450–457. Springer, Heidelberg (2008)
7. Brand, M.: Incremental singular value decomposition of uncertain data with missing values. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2350, pp. 707–720. Springer, Heidelberg (2002)
8. Skočaj, D., Leonardis, A.: Weighted and robust incremental method for subspace learning. In: ICCV, vol. 2, pp. 1494–1501 (2003)
9. Cootes, T.F., Taylor, C.J.: A mixture model for representing shape variation. Image and Vision Computing 17, 567–573 (1999)

10. McLachlan, G.J., Basford, K.E.: Mixture Models: Inference and Applications to Clustering, vol. 84. Marcel Dekker, Inc., New York (1998)
11. Fussenegger, M., Roth, P., Bischof, H., Deriche, R., Pinz, A.: A level set framework using a new incremental, robust Active Shape Model for object segmentation and tracking. Image and Vision Computing 27, 1157–1168 (2009)
12. Cootes, T.F., Taylor, C.J.: Active shape model search using local grey-level models: a quantitive evaluation. In: British Machine Vision Conference, pp. 639–648 (1993)
13. Crappé, O., Moulines, E.: On-line expectation-maximization algorithm for latent data models. Journal of the Royal Statistical Society B 71, 593–613 (2009)

Segmentation of the Infarct and Peri-infarct Zones in Cardiac MR Images

Aditya Gupta and Mubarak Shah

Department of Computer Science, University of Central Florida, USA

Abstract. This paper presents a novel approach for segmentation of the infarct and peri-infarct tissue in the left ventricular wall of the heart. This paper is motivated by a recent finding that shows the infarct and peri-infarct zones to be independent predictors of post myocardial infarction. This paper proposes a method to segment the endocardial and epicardial contours of the left ventricle in the presence of the enhanced infarct and peri-infarct tissues. A level set method using shape priors, obtained from a 3D active appearance model of the ventricle wall on cine MR images is presented. From the extracted 3D cardiac ventricular wall, a method is proposed to segment the infarct and peri-infarct tissues using intensity, volume, shape and heart wall thickness features. The parameters of end-diastolic volume, end-systolic volume, myocardial mass, ejection fraction and infarct and peri-infarct mass are computed using the proposed method and compared with the gold standard provided by the cardiologists. Promising results and comparisons demonstrate the potential of our approach for a practical computer assisted diagnostic system.

1 Introduction

Over the past decade, the use of image processing methods to quantitatively analyze the acquired image data has developed rapidly. Methods have evolved to quantify parameters crucial to diagnose heart diseases such as coronary artery disease using cardiac MR imaging. Particularly, there is interest in the assessment of volumetric change of the infarct and the peri-infarct tissue in the ventricular wall to monitor myocardial infarction (MI). This is based on recent studies that have shown that the infarct and peri-infarct zone is an independent predictor of post MI mortality [1,2]. Animal and human studies have demonstrated that this technique delineates infarct morphology with a high degree of correlation to infarct morphology by pathologic analysis [3]. Furthermore, cardiac MR performed in a prospective outcome study of 100 patients with coronary artery disease has shown that infarct mass is an independent predictor of mortality [4].

Consistent with these results, other cardiac MR investigators have also performed similar studies [5,6]. In addition to infarct size the peri-infarct zone or border zone or gray zone has generated significant clinical interest. Yan, et al [2] showed that the peri-infarct border zone may be an important arrhythmogenic substrate. They utilized a computer-assisted, semiautomatic algorithm to quantify the total infarct size and divided it into the core and peri-infarct regions

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 31–41, 2010.

based on signal-intensity thresholds (>3 standard deviation and 2 to 3 standard deviation above remote normal myocardium, respectively). The peri-infarct zone was normalized as a percentage of the total infarct size. The investigator concluded that for patients with a prior myocardial infarction, the extent of the peri-infarct zone characterized by cardiac MR provides incremental prognostic value beyond left ventricular systolic volume index or ejection fraction.

In another study, Schmidt [7] used gadolinium-enhanced images for infarct characterization. The investigator pre-specified the definitions of 2 standard intensity thresholds that would distinguish the dense, infarct core from the heterogeneous infarct periphery and applied them to the study group. They also used a simplified version of the full-width half-maximum method [8] to define the infarct "core." After the endocardial and epicardial borders were traced by a trained observer, the myocardial segment containing the region of high standard intensity myocardium was outlined, and the maximum standard intensity within this region was determined. The infarct core was then defined as myocardium with standard intensity >50 % of the maximal intensity [8].

Despite the interest in the infarct and peri-infarct tissues, the ability to quantitatively analyze these tissues from the acquired images is still not sufficiently available in routine clinical care. Infarct size measures by human manual contouring and by computerized simple intensity thresholding based on the standard deviation of normal myocardial signal intensities have been shown to significantly overestimate the infarct area in a laboratory animal model [9]. Thus in this paper, we propose a method to segment the infarct and the peri-infarct tissue from the left ventricular wall in an automatic manner. The first step to achieve this segmentation is to determine the endocardial and epicardial heart contours in the presence of the enhanced infarct and peri-infarct tissues. There is extensive research on the segmentation of the endocardial and epicardial walls [10–13] using techniques like deformable models and active contours (snakes) based on the edge, shape and intensity information. But none of these techniques analyze the problem in the presence of infarct and peri-infarct tissue. The delay enhanced infarct and peri-infarct tissues have a similar intensity as the blood in the ventricle in delay enhanced MR images. This makes the problem of left ventricular wall segmentation even more challenging. We propose to achieve this segmentation by a two step method. The first method accomplishes the segmentation of the ventricular wall on short axis cine MR images using 3D active appearance model. In the second step, we propose to use the contours extracted from the 3D AAM in a level set technique based on shape priors. After the 3D left ventricular wall is extracted, we propose a set of features based on intensity, volume and wall thickness to segment the infarct and peri-infarct tissues.

2 Left Ventricle Segmentation

This section discusses the method proposed for segmenting the endocardial and epicardial contours of the left ventricle based on the development of 3D Active appearance model [14]. The Active appearance model (AAM) is a statistical

approach that exploits a priori knowledge of the cardiac shape and image appearance from expert-segmented training examples. In our method, the AAM model is trained over short axis CINE cardiac MR images as in these images the infarct and peri-infarct tissue have a similar intensity as the normal ventricular wall. Prior to implementing the AAM, the cine MR images are subjected to background substraction to extract the cardiac region for improved performance of the model. The CINE MR dataset consists of different slices of the heart taken along the long axis of the subjects at different time instants from end diastole to end systole to next end diastole. Since these images show the ribs, lungs and other structures around the heart, we extract the heart in these images by identifying the pixels in the current frame that deviate significantly from the background. A common approach is the approximated median background substraction method. In this, the running estimate of the median is incremented by one if the input pixel is larger than the estimate, and decreased by one if smaller. This estimate eventually converges to the median. The largest connected component from the foreground identifies the cardiac region.

2.1 Development of 3D Active Appearance Model

The statistical method of 3D AAM has been previously used for the segmentation of the left ventricular wall [11] [10]. Hence in this paper, the implementation of 3D AAM is very briefly discussed. The development of the 3D shape model is achieved by a unique sampling of the left ventricular surface, alignment of the shape samples using Procrustes analysis and principal component analysis (PCA) on the 3D models. The 3D appearance model is developed by eliminating the shape variation using piecewise affine warping on a tetrahedral representation of volume using the 3D Delaunay triangulation algorithm. After the warping phase, the images are intensity normalized and PCA is performed on the training data to compute the 3D appearance model. Now the shape information (shape vectors) and the intensity information (intensity vectors) are combined into a single vector and a final PCA gives the 3D AAM. The model is superimposed over known annotated data and the model parameters are disturbed slightly. The affine transformation, intensity parameters and the appearance co-efficients are perturbed from their original value. The resulting difference between the model and image is determined. A relation between image differences and model parameter differences is estimated by multivariate linear regression.

To segment the 3D image the model has to be matched to it. This matching will be achieved by iteratively minimizing the root-mean-square difference between the model and the image by modifying the model appearance parameters and the affine transformation. This classical optimization problem can be solved efficiently by iteratively estimating the model parameter updates. When the model is matched to unknown data, the model is placed somewhere close to the object which is to be segmented. Then the texture differences are determined and the estimated relation is used to estimate the optimal change of model parameters. The parameter updates are repeated until the texture differences falls below some threshold.

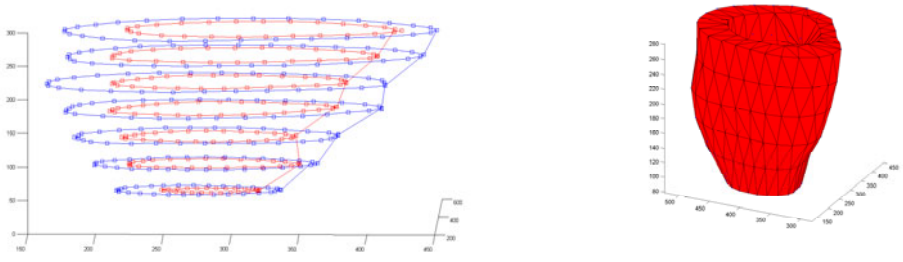


Fig. 1. Figure showing an example of unique sampling of the left ventricular wall and the tetrahedral representation of ventricle

2.2 Segmentation of the Left Ventricular Walls in the Presence of Infarct

The 3D AAM model shows a poor performance when implemented on the single shot delay enhanced MR images as the enhanced infarct and peri-infarct tissues have a similar intensity as the blood in the ventricle and in many cases extend from the endocardium to the epicardium wall. To address this problem, we propose to use a level set based segmentation method that incorporates prior shape information obtained from the 3D AAM. A level set variant of the active shape model has the ability to account for local image features while being able to introduce prior shape knowledge and has been previously implemented in [15]. The endocardial and epicardial contours extracted using the 3D AAM model is used as the prior shape information in the level set space. For the level set representation, consider a closed evolving surface $[C : [0, 1] \rightarrow R^2, p \rightarrow C(p)]$ and let $C(p, t)$ represent the entire surface driven by the propagation of an initial curve $C_0(p)$ according to:

$$C_t(p) = F(p)N(p), C(p, 0) = C_0(p), \tag{1}$$

where F is a scalar function and N is the inward normal. The entire surface is evolved to minimize a metric defined by the curvature and image gradient. The zero-level set function ($\phi = 0$) of a surface z is $z = (x, y, \phi(x, y, t)) \in R^3$. The motion of the surface can be obtained by deriving $\phi(x, y, t) = 0$ with respect to time and space.

$$\phi(C_0(p), 0) = 0, \phi_t(p) = F(p)|\nabla\phi(p)|, \tag{2}$$

where $|\nabla\phi|$ is the norm of the gradient. Let the level set representation evolving over time $\phi : \Omega \times R^+ \rightarrow R^+$ be a Lipchitz function given by:

$$\phi(x, y, t) = \begin{cases} 0, (x, y) \in C(t); \\ +D((x, y), C(t)) > 0, (x, y) \in C_{in}(t); \\ -D((x, y), C(t)) < 0, (x, y) \in C_{out}(t) = [\Omega - C_{in}(t)]. \end{cases} \tag{3}$$

where $(x, y) = p$, $C_{in}(t)$ is the area enclosed by the curve C, $D((x, y), C(t))$ is the minimum Euclidean distance between the pixel (x, y) and $C(t)$ at time t.

The Dirac (δ) and Heaviside functions can be used to define terms along C and the interior and exterior of C:

$$\begin{array}{l}(x,y) \in \Omega : \{lim_{\alpha \to 0_+}[\delta_\alpha(\phi(x,y))] = 1\} = C \\ (x,y) \in \Omega : \{lim_{\alpha \to 0_+}[H_\alpha(\phi(x,y))] = 1\} = C_{in}.\end{array} \tag{4}$$

Using the above level set representation, the curve propagates with respect to known shape properties derived from the left ventricular contours.

Let us consider a set C_i of contours obtained from 3D AAM. Typically for a patient, this consists of 200 to 300 cine short axis MR images along the axis of the heart from end-diastole to end-systole to next end-diastole. The 2D active shape model is constructed to get an equation in the form of the level set function ϕ:

$$\phi = \phi_M + \sum_{j=1}^{m} p_j U_j, \tag{5}$$

where m is the number of modes of variation, U_j are the eigenvectors from PCA and p_j are linear weight factors within the allowable range defined by the eigenvalues.

There is an ideal transformation $A = (A_x, A_y)$ between the shape prior and the propagating level set representation ϕ. The optimal transformation will satisfy the conditions:

$$(x,y) \to A(x,y)\phi(x,y) \approx \phi_M(A(x,y)), \forall (x,y) \in \Omega \tag{6}$$

The sum of squared differences is used for optimization. Scale variation is added to the transformation $A = (s, \theta, T_x, T_y)$. Estimating the prior in the vicinity of the zero crossing and close to the origin of transformation, the energy minimization functional is:

$$E(\phi, A) = \iint_{\Omega} \delta_\epsilon(\phi)(s\phi - \phi_M(A))^2 \, d\Omega \tag{7}$$

By calculus of variations, the equation of evolution for ϕ is given by:

$$\frac{d\phi}{dt} = -2\delta_\epsilon(\phi)s(s\phi - \phi_M(A)) \tag{8}$$

To map ϕ to the best ϕ_M the parameters of transformation are also updated according the equations:

$$\begin{cases} \frac{ds}{dt} = 2\iint_{\Omega} \delta_\epsilon(\phi)(s\phi - \phi_M(A))(-\phi + \nabla\phi_M(A).\frac{\partial A}{\partial S}\, d\Omega; \\ \frac{d\theta}{dt} = 2\iint_{\Omega} \delta_\epsilon(\phi)(s\phi - \phi_M(A))(\nabla\phi_M(A).\frac{\partial A}{\partial \theta}\, d\Omega; \\ \frac{d}{dt}\begin{matrix} Tx \\ Ty \end{matrix} = 2\iint_{\Omega} \delta_\epsilon(\phi)(s\phi - \phi_M(A))(\nabla\phi_M(A).\frac{\partial A}{\partial \begin{matrix} Tx \\ Ty \end{matrix}}\, d\Omega; \end{cases} \tag{9}$$

The ideal transformation derived from the equations will map each value of current representation to the best level set representation belonging to the class of the training shapes. Typically, pixels of infarct have intensity larger than or equal to 3 standard deviations (SD) of a normal myocardial region and pixels of peri-infarct have intensity between 2 SDs and 3 SDs [17].

2.3 Segmentation of the Infarct and Peri-Infarct Tissue

From the extracted ventricular wall, the mass and volume of the infarct and the peri-infarct tissues can be determined. Due to the enhancement of these tissues, the left ventricular wall has primarily two intensities - the dark normal ventricular tissue and the bright infarct and peri-infarct tissues. A simple intensity histogram of the ventricular wall separates these regions. But segmentation based on an intensity based model significantly overestimates the infarct area [9]. Even a slight error in ventricular wall segmentation leads to the possibility of the blood in the ventricular wall to be identified as infarct. Hence we propose to use certain features that uniquely identify the infarct and peri-infarct tissues. The intensity based potential infarct regions are analyzed for their volume and shape. The nature of the infarct and the peri-infarct tissues is that they originate from the endocardial surface and spread towards the epicardium. Hence, if the volume of any of the regions is below a certain threshold or if any of the regions appear on the epicardial wall and are not connected to the endocardial wall then these regions are identified as false detection. In previous studies on animals and humans [16, 17], the change in the myocardial wall motion and wall thickening for myocardial infarction is clinically determined. Particularly in [16], the relation of wall thickening and motion for infarcted tissue is determined using two dimensional echocardiography. Comparing infarcted with normal zones in each slice, the percentage wall thickening shows a clear separation with little overlap. Regional percentage of systolic thinning (Systhin) is calculated as:

$$Systhin = \frac{Th_{ES} - Th_{ED}}{Th_{ED}} \times 100 \tag{10}$$

where Th_{ES} is the thickness of the myocardial segment at end-systole and Th_{ED} is the thickness of the segment at end-diastole (in cm). Negative values indicate systolic wall thinning. The unique sampling of the ventricular surface is achieved in a method described in 2.2. At each of the points the regional percentage of systolic thinning is computed. If any of the potential infarct regions have a significant positive Systhin value, then these regions are identified as false detections. From the clinical study in [16] it is also determined that the wall motion feature may over-estimate the infarct zone. Since we use these features on the regions segmented using intensity histogram, the problem of over-estimation of the infarct zone should not arise.

After implementing the features described above, the infarct and the peri-infarct tissues are segmented using k-means algorithm. The surface areas and the mass of the infarct and peri-infarct regions can then be determined. The number of pixels identified as infarct (Inf_pixel) and peri-infarct (Pinf_pixel) in each slice is identified. The parameters of in-plane resolution (ip_res) and effective slice thickness (thickness) are known from the MR examination. The infarct volume (Inf_vol) and mass (Inf_mass) is computed on the individual slices using the equation:

$$\begin{aligned} Inf\_vol &= Inf\_pixel \times ip\_res \times thickness \\ Inf\_mass &= Inf\_vol \times 1.05 \end{aligned} \tag{11}$$

The peri-infarct mass and volume is computed using similar equations. The values obtained are compared with the values obtained from manual contouring and the results are discussed in the next section.

3 Results

In this section, we present our results for (i) the segmentation of left ventricle using 3D AAM on short-axis cine images, (ii) segmentation of the left ventricle using the level set method with shape prior on single shot delay enhanced images, and (iii) segmentation of the infarct and the peri-infarct tissue. Cardiac MR images consisting of the cine short axis image sequences and the single shot delay enhanced images were collected for 25 patients with an ejection fraction <50%. The cine short axis image sequence of each patient consists of exactly 25 frames and the number of slices acquired along the long axis of the subjects ranged between 8 and 12. Each patient's dataset consists of 200 to 300 2D images, with a resolution of 256 X 256 pixels. Background modeling on the cine images of each patient results in images of resolution 125 X 125 pixels. Of the 25 patients, five of them had a clearly visible infarct tissue. Each patient's single shot delay enhanced dataset consists of 10 images acquired along the long axis. The 3D AAM was trained by manually extracting the endocardial and epicardial contours from the cine images of 10 patients. The model was tested on the cine images of the remaining 15 patients. The endocardial and epicardial contours extracted with the 3D AAM showed good agreement with the manual contours on majority of the slices except on the extreme apex slices. Manual contours were present in the extreme apical slices but the 3D AAM failed to identify contours. Parameters of end diastolic volume (EDV), end systolic volume (ESV), myocardial mass (MM) and ejection fraction (EF) are computed for manual contouring (gold standard) technique and the 3D AAM using the equations:

$$\begin{aligned} EDV &= (\text{Summation of pixel areas in diastole}) \times (ip\_res) \times (thickness) \\ ESV &= (\text{Summation of pixel areas in systole}) \times (ip\_res) \times (thickness) \\ MM &= (\text{Epicardial volume - Endocardial volume}) \text{ in diastole} \times 1.05 \\ EF &= (EDV - ESV)/EDV \end{aligned} \tag{12}$$

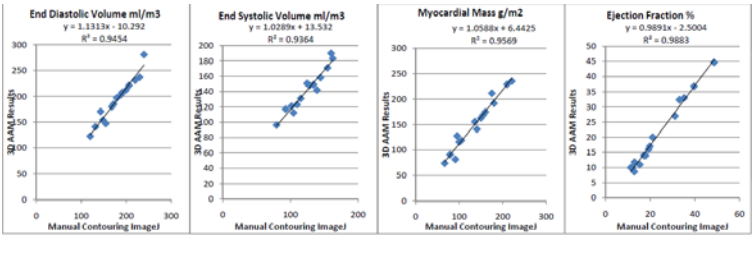


Fig. 2. Regression plots comparing the performance of 3D AAM with manual contouring in cine MR images

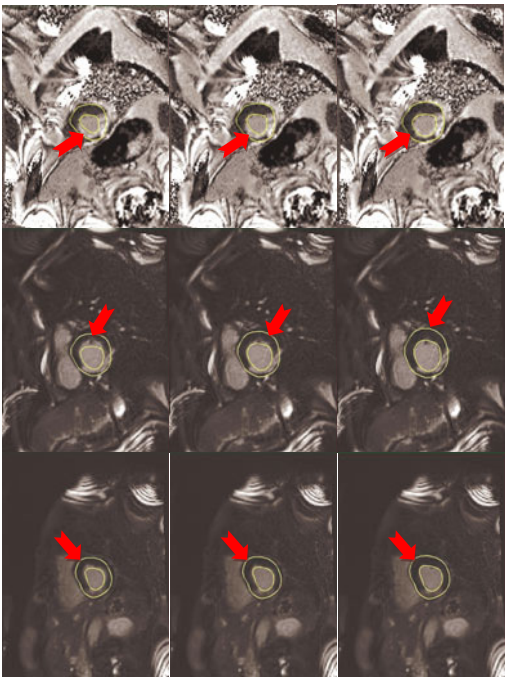


Fig. 3. The performance of the propagating level set function on three patients. The images on the left show the initial contours and the images to the right show the final output.

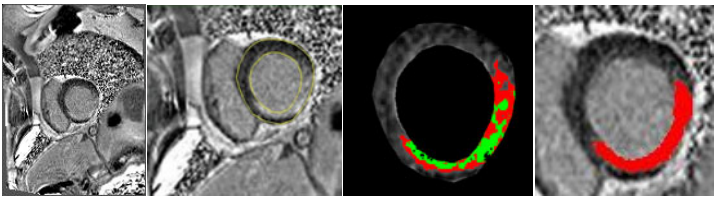


Fig. 4. a) Original Delay enhanced SAX MR images for 3 patients b)Result of ventricular wall segmentation using level set with shape prior on background modeled images c)Segmentation of the infarct (green) and peri-infarct tissue (red) using the proposed features and k-means d)Manual contouring of the infarct tissue using ImageJ

The regression plots comparing the performance of 3D AAM with manual contouring for the above mentioned parameters is shown in figure 2. Correlation coefficient of 94.54%, 93.64%, 95.69% and 98.83% is obtained for the end diastolic volume, end systolic volume, myocardial mass and ejection fraction respectively.

For segmenting the endo and epicardial contours on the single shot delay enhanced images for the 5 patients with clearly visible infarct tissue, an active shape model is developed from the endocardial and epicardial slices obtained

	Manual Contouring - ImageJ					Automatic Method using Features				
	Infarct Mass in grams					Infarct Mass in grams				
	Patient#1	Patient#2	Patient#3	Patient#4	Patient#5	Patient#1	Patient#2	Patient#3	Patient#4	Patient#5
Slice 1	0	0	0	0	2.801446	0	0	0	0	1.939463
Slice 2	0	0	0	0	4.870207	0	0	0	0	3.878926
Slice 3	0	3.491033	0	0	1.896364	0	5.042603	0	0	1.853264
Slice 4	2.542851	9.438719	4.137521	9.654215	0	3.534132	9.309421	5.128802	12.75736	0
Slice 5	5.171901	9.438719	6.982066	8.490537	0	4.913306	11.59368	9.955909	10.4731	0
Slice 6	2.973843	7.887149	0	9.438719	0	3.404835	10.5162	0	8.749132	0
Slice 7	0	3.620331	0	7.930248	0	0	4.956405	0	5.775289	0
Slice 8	0	0	0	7.369959	0	0	1.896364	0	9.223223	0
Slice 9	0	0	0	9.137025	0	0	0	0	10.90409	0
Slice 10	0	0	0	2.887645	0	0	0	0	4.654711	0

Fig. 5. Infarct mass computation on 10 slices of 5 patients for manual contour method and the proposed feature based method

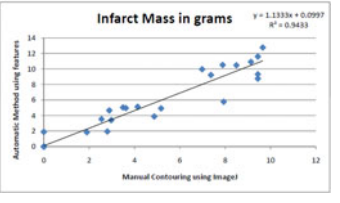


Fig. 6. Regression plots comparing the performance of 3D AAM with manual contouring in cine MR images

from the 3D AAM model for each patient. The single shot delay enhanced images are aligned to the cine images using rigid registration. The end diastolic contours of the closest corresponding slice along the long axis of each patient is used as the starting contour for the level set method. The active shape model for the patient is used as the shape prior for the propagating level set. The result of the proposed level set method is shown in 3. The images on the left show the starting contour for the level set functions for three patients which are obtained from 3D AAM at end diastole. An intermediate stage of the propagating contour is shown in the middle images and the final output is shown in the rightmost images. Particularly, the region marked with the red arrows in the images from left to right clearly shows the improvement in performance of the propagating level set.

The complete results of background modeling, the level set method and the segmentation of the infarct and the peri-infarct tissues for a particular slice of a patient is shown in 4. The figure also shows the comparison of the infarct and peri-infarct tissues segmented with the manually contoured infarct tissue. Promising results demonstrate that the proposed method can be used for automatic segmentation of the endocardial and epicardial contours in the presence of the infarct and peri-infarct tissue. The infarct mass is computed on the 10 slices of the 5 patients. Figure 5 shows the comparison of the infarct mass (in grams) computed using equation 11, for manual contouring (using ImageJ) method and the proposed method. Regression plot in figure 6, shows a high correlation

co-efficient of 94.33%. The comparison for the segmented peri-infarct tissue with manual contouring is not shown as this zone is difficult to identify manually (using ImageJ).

4 Conclusion

In this paper, we have proposed a method of segmenting the endocardial and epicardial contours of the left ventricle in cardiac MR images in the presence of the enhanced infarct and peri-infarct tissues. A level set function was formulated which includes the prior shape information obtained from an active shape model. Since we propose to develop the active shape model as a patient specific model, promising segmentation results are obtained. Using these results, we have proposed to use a unique set of features for segmenting the infarct and the peri-infarct tissues. High correlation co-efficients of above 90% are obtained for all the segmentation results. In our future work, we will explore including temporal information in the 3D AAM model and additional features like wall motion to identify infarct and peri-infarct tissues with higher accuracy. We believe that incorporating these features and using a larger training dataset will lead to a better segmentation performance.

References

1. Bello, D., Fieno, D., Goldberger, J., Kadish, A.: Infarct size is an independent predictor of mortality in patients with coronary artery disease. American College of Cardiology 45(3A), 288A (2005)
2. Yan, A., Shayne, A., Brown, K., Gupta, S., Chan, C., Luu, T., Carli, M.D., Reynolds, G., Stevenson, W., Kwong, R.: Characterization of the peri-infarct zone by contrast-enhanced cardiac magnetic resonance imaging is a powerful predictor of post-myocardial infarction mortality. Circulation 4(114) (2006)
3. Bello, D., Shah, D., Farah, G., Luzio, S.D., Parker, M., Johnson, M., Cotts, W., Klocke, F., Bonow, R., Judd, R., Gheorghiade, M., Kim, R.: Gadolinium cardiovascular magnetic resonance predicts reversible myocardial dysfunction and remodeling in patients with heart failure undergoing beta-blocker therapy. Circulation 108 (2003)
4. Bello, D., Einhorn, A., Fieno, D., Goldberger, J., Kadish, A.: Cardiac magnetic resonance imaging: Infarct size is an independent predictor of mortality in patients with coronary artery disease. Circulation (2010) (in revision)
5. Roes, S., Kelle, S., Kaandorp, T., Kokocinski, T., Poldermans, D., Lamb, H., Boersma, E., van der Wall, E., Fleck, E., de Roos, A., Nagel, E., Bax, J.: Comparison of myocardial infarct size assessed with contrast-enhanced magnetic resonance imaging and left ventricular function and volumes to predict mortality in patients with healed myocardial infarction. American Journal of Cardiology 100, 930–936 (2007)
6. Wu, E., Ortiz, J.T., Tejedor, P.: Infarct size by contrast enhanced cardiac magnetic resonance is a stronger predictor of outcomes than left ventricular ejection fraction or end-systolic volume index: prospective cohort study. Heart 94, 730–736 (2008)

7. Schmidt, A., Azevedo, C., Cheng, A., Gupta, S., Bluemke, D., Foo, T., Gerstenblith, G., Weiss, R., Marbn, E., Tomaselli, G., Lima, J., Wu, K.: Infarct tissue heterogeneity by magnetic resonance imaging identifies enhanced cardiac arrhythmia susceptibility in patients with left ventricular dysfunction. Circulation 115, 2006–2014 (2007)
8. Simonetti, O., Kim, R., Fieno, D., Hillenbrand, H., Wu, E., Bundy, J., Finn, P., Judd, R.: An improved MRI technique for visualization of myocardial infarction. Radiology 218, 215–223 (2001)
9. Hsu, L.Y., Ingkanisorn, P., Kellman, P., Aletras, A., Arai, A.: Quantitative myocardial infarction on delayed enhancement MRI. part i: Animal validation of an automated feature analysis and combined thresholding infarct sizing algorithm. Journal of Magnetic Resonance Imaging 23, 298–308 (2006)
10. Andreopoulos, A., Tsotsos, J.: Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI. Medical Image Analysis 12(3) (2008)
11. Mitchell, S., Bosch, J., Lelieveldt, B., van der Geest, R., Reiber, J., Sonka, M.: 3-d active appearance models: segmentation of cardiac MR and ultrasound images. IEEE Transactions on Medical Imaging 21(9) (2002)
12. Lorenzo-Valds, M., Sanchez-Ortiza, G., Elkingtonb, A., Mohiaddinb, R., Rueckert, D.: Segmentation of 4d cardiac MR images using a probabilistic atlas and the em algorithm. Medical Image Analysis 8(3) (2004)
13. Kausa, M., von Berga, J., Weesea, J., Niessenb, W., Pekar, V.: Automated segmentation of the left ventricle in cardiac MRI. Medical Image Analysis 8(3) (2004)
14. Cootes, T., Edwards, G., Taylor, C.: Active appearance models. IEEE PAMI 23(6) (2001)
15. Leventon, M., Grimson, E., Faugeras, O.: Statistical shape influence in geodesic active contours. IEEE Conf. on Comp. Vision and Patt. Recog. 1, 316–323 (2000)
16. Lieberman, A., Weiss, J., Jugdutt, B., Becker, L., Bulkley, B., Garrison, J., Hutchins, G., Kallman, C., Weisfeldt, M.: Two-dimensional echocardiography and infarct size: Relationship of regional wall motion and thickening to the extent of myocardial infarction in the dog. Circulation 63(4) (1981)
17. Gibson, D., Prewitt, T., Brown, D.: Analysis of left ventricular wall movement during isovolumic relaxation and its relation to coronary artery disease. Heart 38, 1010–1019 (1976)

Spatial-temporal Constraint for Segmentation of Serial Infant Brain MR Images

Feng Shi[1], Pew-Thian Yap[1], John H. Gilmore[2], Weili Lin[3], and Dinggang Shen[1]

[1] IDEA Lab, [3] MRI Lab, Department of Radiology and BRIC,
University of North Carolina at Chapel Hill, USA
[2] Department of Psychiatry, University of North Carolina at Chapel Hill, USA
dgshen@med.unc.edu

Abstract. Longitudinal infant studies offer a unique opportunity for revealing the dynamics of rapid human brain development in the first year of life. To this end, it is important to develop tissue segmentation and registration techniques for facilitating the detection of global and local morphological changes of brain structures in an infant population. However, there are two inherent challenges involved in development of such techniques. *First*, the MR images of the isointense stage – the duration between infantile and early adult stages in the first year of life – have low gray-white matter contrast. *Second*, temporal consistency cannot be preserved if segmentation and registration are performed separately for different time-points. In this paper, we proposed a 4D joint registration and segmentation framework for serial infant brain MR images. Specifically, a spatial-temporal constraint is formulated to make optimal use of T1 and T2 images, as well as adaptively propagate prior probability maps among time-points. In this process, 4D registration is employed to determine anatomical correspondence across time-points, and also a multi-channel segmentation algorithm, guided by spatial-temporally constrained prior tissue probability maps, is applied to segment the T1 and T2 images simultaneously at each time-point. Registration and segmentation are iterated as an Expectation-Maximization (EM) process until convergence. The infant segmentations yielded by the proposed method show high agreement with the results given by a manual rater and outperform the results when no temporal information is considered.

1 Introduction

Infants undergo rapid developments of brain structures and functions after birth, especially in the first year of life [1]. Tissue segmentation, a standard procedure in image analysis, is however challenging for infant images due to their inherently insufficient spatial resolution and low tissue contrast. Recently, numerous methods have been proposed for automatic brain segmentation of neonates (less than 3 months) [2-4] and infants (over 1-year-old) with varied success. However, little attention is paid to the segmentation of infants in the duration between birth and one year of age. Brain in this age range is affected by a combination of factors such as maturation and myelination. We show a series of longitudinal MR images in Fig. 1 for an infant scanned at an approximately 3-months interval, starting from two weeks.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 42–50, 2010.

We can observe that, in the T1 images, white matter (WM) initially has lower signal intensity than gray matter (GM) (e.g., 2 weeks), but eventually becomes higher and reaches the adult-like pattern after 9 months. A reversed contrast but same pattern can be observed in the T2 images. Our observations agree with Dietrich *et al.*'s findings [5], showing that three distinct gray-white matter signal intensity patterns appears in chronological order in images of developmentally normal infant – infantile (birth), isointense, and early adult (10 months onward). Images at isointense stage have the lowest gray-white matter contrast and have not been fully explored in previous studies.

To address this problem, we first make two observations. *First*, for the purpose of segmentation, T2 images are usually preferred in the infantile stage and T1 images in the early adult stage. This suggests that balance is needed between T1 and T2 images for guiding the segmentation. *Second*, infantile and early adult stage images can be segmented relatively easily compared to those from the isointense stage. This hints that priors could be constructed from both two stages and propagated to the isointesnse images for segmentation guidance.

Temporal consistency is an important consideration factor in longitudinal infant studies. To analyze a 4D dataset, a conventional pipeline would perform segmentation on the images separately for each time-point. The segmented images are then registered across time-points for determining anatomical correspondences. Tissue density maps characterizing the volume and shape changes are finally employed for comparing brain changes over time. Great successes using this approach have been achieved in understanding early brain development [1]. However, inconsistency might be introduced by performing segmentation and registration independently with disregard of the temporal constraint, which could in the end reduce the reliability and statistical power of the findings.

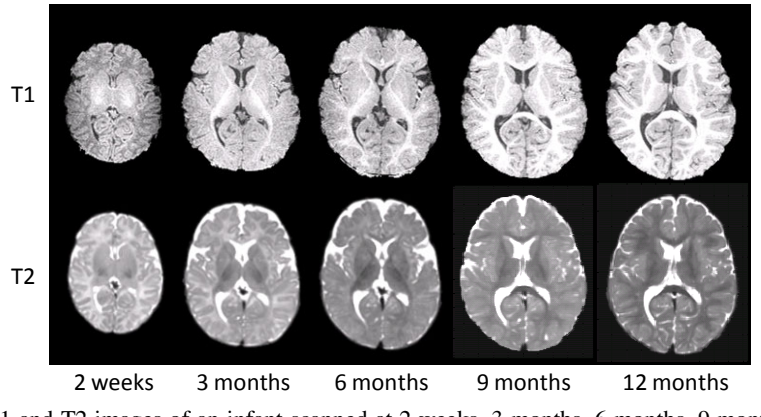


Fig. 1. T1 and T2 images of an infant scanned at 2 weeks, 3 months, 6 months, 9 months, and 12 months. The gray-white matter contrast varies as a function of time, with three major stages – infantile, isointense, and early adult.

In this paper, we address the above-mentioned issues and propose a 4D joint registration and segmentation framework, the key idea of which is illustrated in Fig.2. Segmentation of images from the isointense stage is difficult, but this difficulty can be

ameliorated if we borrow information from neighboring time-points, providing constraint from temporal correspondences. By taking information from all time-points into consideration, temporal consistency can also be better preserved. To achieve this aim, we formulate a spatial-temporal constraint to make optimal use of the T1 and T2 images and to adaptively propagate the prior probability maps of each time-point to its neighboring time-points. In this process, 4D registration is employed to determine anatomical correspondences across time-points, and a multi-channel segmentation algorithm, which take into consideration the spatial-temporal constraint as prior, is used for segmenting the image at each time-point. Registration and segmentation are iterated as an Expectation-Maximization (EM) process until convergence. Details of the proposed method are given in the following sections. Note that although detection of within-tissue differences caused by myelination and maturation is also an important direction in early development study, this paper focuses on segmenting brain into general GM, WM, and CSF, in which WM contains both myelinated and unmyelinated tissue, GM contains both the cortical GM and basal ganglia.

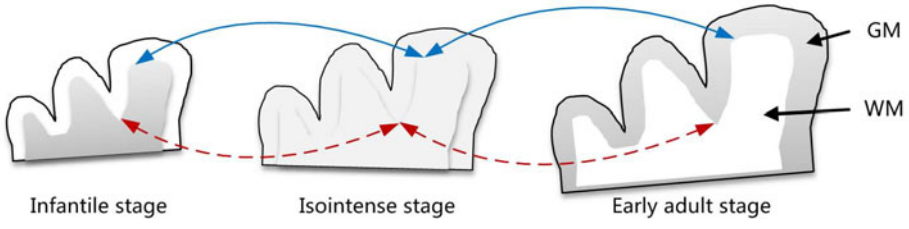


Fig. 2. Spatial-temporal constraint for joint registration and segmentation of serial images from the first year of life. Three major gray-white matter patterns are shown. Probabilistic maps of tissues from the images of infantile and early adult stages are used to guide the segmentation of brain tissues in the isointense stage.

2 Method

We propose a 4D joint registration and segmentation framework, which alternates between 4D registration, spatial-temporal constraint construction (i.e., prior probability maps for each time-point), and multi-channel segmentation. T1 and T2 MR images are collected from 5 time-points – 2 weeks, 3 months, 6 months, 9 months, and 12 months. Images are skull-stripped with BSE [6], followed by manual editing to ensure the clean removal. Intensity inhomogeneity is corrected with N3 [7]. For initialization, we first segment the last time-point T1 image with an adaptive K-means algorithm. The resulting segmentation is taken as a longitudinal prior for guiding the initial segmentation of earlier time-point images independently, using the method described in [4]. This preliminary segmentation is fed as an input to our framework. In the following subsections, we will discuss the formulation of the spatial-temporal constraint, followed by the 4D joint registration and segmentation mechanism.

2.1 Formulation of Spatial-temporal Constraint

T1-T2 Preference Factor. As mentioned above, T2 is preferred for segmenting infantile images, and T1 for images from the early adult stage. To achieve balance and make optimal use of T1 and T2 images, we define *preference factors* $\omega_{T_1}(t)$ and $\omega_{T_2}(t)$ for each time-point t, where $\omega_{T_1}(t) + \omega_{T_2}(t) = 1$, $t = \{1,2,3,4,5\}$, for reflecting the reliance of the segmentation algorithm on the T1 and T2 images, respectively. Since segmentation reliability is dependent on gray-white matter contrast, we compute the preference factors by comparing the histograms of GM and WM using the Jensen–Shannon (JS) divergence, which measures the difference of the two probability distributions and is symmetric. Initial GM and WM intensity distributions are obtained with the help of the initial segmented images and are updated as we have a more segmentation as we iterate through the registration-segmentation process. A plot of the JS values of the subject shown in Fig. 1 is given in Fig. 3(a). It can be observed that the tissue contrast of the T1 image is initially low and eventually increases to a value higher than that of the T2 image. The T2 image has a relatively good contrast throughout, except during the isointense stage. We thus define the *preference factors* as below:

$$\omega_{T_1}(t) = JS(t, T_{p1}) / (JS(t, T_1) + JS(t, T_2)) \tag{1}$$

$$\omega_{T_2}(t) = 1 - \omega_{T_1}(t) \tag{2}$$

where $JS(t, T)$ is the JS divergence value, as a function of time-point t and image modality $T = \{T_1, T_2\}$. These preference factors will be employed in the multi-channel segmentation algorithm for utilizing the T1 and T2 images more effectively.

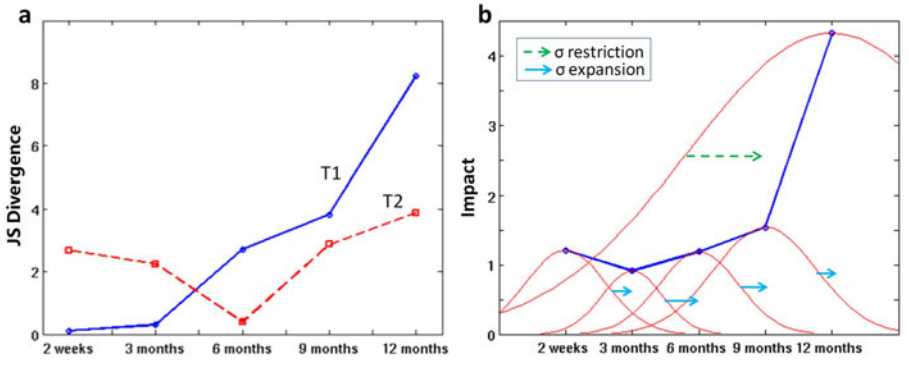


Fig. 3. (a) GM-WM JS divergence values signifying tissue contrasts in the T1 and T2 images of 5 different time-points. (b) Temporal impact for each time-point, and the Gaussian distribution of the impact factor for each time-point and their evaluation direction.

Temporal Impact Factor. We combine the T1 and T2 preference factors to obtain a measure called the *temporal impact factor*:

$$\omega(t) = \omega_{T_1}(t) * JS(t,T_1) + \omega_{T_2}(t) * JS(t,T_2)$$
$$= \frac{\left(JS(t,T_1)\right)^2 + \left(JS(t,T_2)\right)^2}{JS(t,T_1) + JS(t,T_2)} \tag{3}$$

where the impact factor $\omega(t)$ is weighted as a Contraharmonic mean of JS values of T1 and T2 images, which assigns more weight on the one with larger JS value.

This measure describes the across-time-point influence of an image. Time-point with high contrast image has a high impact value, implying that its influence will be propagated a long distance in the temporal dimension, and its probability map would have a greater impact on guiding the segmentation of other time-points. An example of this is the time-point of 12 months (see Fig. 3(b)). Due to the temporal changes in the brain, such as the generation of synapses and myelination, this impact is expected to be lower when propagated further along the temporal dimension. To simulate this, we incorporate the temporal impact factor into a Gaussian kernel and define:

$$G_{t,\sigma}(\tau) = \omega(t) exp\left(-\frac{(\tau - t)^2}{2\sigma^2}\right) \tag{4}$$

where $(\tau - t)$ is the temporal distance between the current time-point t and time-point τ (with a unit temporal distance defined as equivalent to 3 months), and σ is a variance parameter for controlling the flatness of the Gaussian distribution. The parameter σ starts with $\sigma = k\omega(t)$ and ends with a value of 1. k is a factor dealing with the scale difference of the temporal and the impact axes. $G_{t,\sigma}(\tau)$ are indicated by red curves in Fig. 3(b).

The segmentation prior probabilistic map at a certain time-point t, is obtained via a linear combination of the segmentation probabilistic map at each time-point, weighted by the temporal impact factor:

$$p_t(L) = \frac{\sum_{\tau=1}^{N} G_{\tau,\sigma}(t) \times p_\tau(L|D)}{\sum_{\tau=1}^{N} G_{\tau,\sigma}(t)} \tag{5}$$

where L is a segmentation label, D is the deformation field pointing from time-point τ to t, which is used to warp the probabilistic map from the space of time-point τ to the space of time-point t, resulting in $p_\tau(L|D)$, and $p_t(L)$ is the obtained prior probabilistic map of L at time-point t.

Notice that the segmentation result at each time-point is progressively refined and the dependence of other time-points is gradually reduced. Images from a certain time-point might have a large impact initially by providing significant prior information to other time-point images. As the segmentation result at the influenced time-point is getting better, less information needs to be borrowed from the neighboring time-points, and temporal impact factor of the neighboring time-point is decreased. On the other hand, an initially small temporal impact factor may also be increased, if the segmentation result of the current time-point is getting better, to help share the segmentation information with neighboring time-points. These two temporal impact

factor changing patterns can be controlled by the parameter σ in (4), to expand or to restrict as the temporal impact, as illustrated in Fig. 3(b).

2.2 4D Joint Registration and Segmentation Framework

4D registration helps determine anatomical correspondence between time-points, and also provides a means to warp the segmentation probabilistic maps of a certain time-point to other time-points. Segmentation, on the other hand, helps provides information for better registration of the images at each time-point. This can be described as an Expectation-Maximization (EM) process [8], with the E- and M-steps detailed as follows:

a) **E-step: Segmentation.** Given the intensity images $\{I_{t,T}, T = T_1, T_2\}$ and prior probability maps $p_t(L)$ obtained from M-step, at each time-point *t*, we can obtain a single segmentation S_t for both T1 and T2 images with multi-channel segmentation technique.

b) **M-step: Registration and Constrain Construction.** Deformation D is updated based on the new segmentations $\{S_t\}$ from the E-step, providing anatomical correspondence between time-points. The parameter σ is changed to progressively control the temporal impact factor changing patterns. Then the spatial-temporal constraint $p_t(L)$ is then computed by weighting the warped tissue probability maps $p_t(L|D)$ across time-points, as detailed in subsection 2.1.

These two steps are iterated and progressively refine segmentation and registration. The program stops when segmentation results do not show further significant changes. In what follows, we briefly introduce the 4D registration and multi-channel segmentation approach used in this framework. More details can be found in [8, 9].

4D Registration. Given a group of segmentations $\{S_t\}$ for the input image sets $\{I_{t,T}, T = T_1, T_2\}$, we need to determine the anatomical correspondences across time-points by a series of deformation fields D. A template set $\{S_t^{Template}\}$ is first generated by duplicating the last time-point $(t = 5)$ segmented image (with best contrast) to each other time-point. To register the image sets $\{S_t\}$ to $\{S_t^{Template}\}$, we hierarchically determine the optimal transformation which minimizes the difference between the attribute vectors of images in $\{S_t\}$ and $\{S_t^{Template}\}$ [9]. The attribute vector is composed of the Geometric Moment Invariants (GMI) for extraction of structural information. The cost function for the 4D registrations is:

$$E = E_F + E_B + E_S \tag{6}$$

where E_F and E_B are the attribute vector similarity cost functions related to the forward and backward transforms, and E_S is a spatial-temporal smoothness constraint. 4D registration has been evaluated and utilized in many applications such as in the studies of longitudinal changes of hippocampal volumes [9] and myocardial motion estimation [10].

Multi-Channel Segmentation. Having both T1 and T2 images allows us to make optimal use of the information from both modalities for better segmentation. Some multi-channel segmentation algorithms produce two individual results for T1 and T2

images. However, since both T1 and T2 images reflect the same tissue structures, producing a single segmentation result might be more appropriate. To this end, we employ a multi-channel segmentation algorithm which works with joint probability distributions. Denoting I as the intensity image, we view the T1 and T2 images as a single vector image with 2 elements at each voxel. We extend the single-channel Gaussian mixture model (GMM) segmentation [8] to its multi-channel counterpart by using axis-aligned multivariate Gaussians. In particular, for a voxel with vector $X = (x_1, x_2)^T$, we can estimate the mean $\mu_i = (\mu_{i,1}, \mu_{i,2})^T$ and variance $\Sigma_i = [\Sigma_{i,1}, \Sigma_{i,2}]^T$ for the i-th Gaussian for a certain tissue, and the probability of X belongs to this tissue can be defined as $p_i(X) = \frac{1}{\sqrt{2\pi}|\Sigma_i|^{1/2}} exp\left(-\frac{1}{2}(X-\mu_i)^T {\Sigma_i}^{-1}(X-\mu_i)\right)$. The preference factors are taken into account by defining $\omega = \begin{bmatrix} \omega_{T_1} & 0 \\ 0 & \omega_{T_2} \end{bmatrix}$, and incorporating it in the probability, as follows:

$$p_i(X) = \frac{1}{\sqrt{2\pi}|\Sigma_i|^{\frac{1}{2}}} exp\left(-\frac{1}{2}(X-\mu_i)^T \omega^{\mathrm{T}} {\Sigma_i}^{-1} \omega (X-\mu_i)\right) \tag{7}$$

Initial means and variances are estimated based on a set of previously segmented images. Using Bayes rule, posterior probability maps can be obtained for image I, giving the probabilities of each voxel of belonging to gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). We apply the majority rule and assign the label with the maximal probability to each voxel. This newly updated segmentation label can be used to better estimate the parameters and perform another iteration of segmentation for better accuracy.

3 Experimental Results

To validate our proposed 4D joint registration and segmentation method, we apply it to a group of infants, scanned at 2 weeks, 3 months, 6 months, 9 months, and 12 months, with variations of same time-point scans less than 1 week. These healthy subjects were not sedated for MRI. T1 images were acquired using a 3T head-only MR scanner, with 144 sagittal slices at resolution of 1×1×1 mm^3. Meanwhile, T2 images of 64 axial slices were obtained at resolution of 1.25×1.25×1.95 mm^3. Standard preprocessing steps such as skull stripping and bias correction are performed. T2 images are affine aligned to their T1 counterparts and are resampled with a 1×1×1 mm^3 resolution before further processing. We evaluate the segmentation accuracy on all time points by visual inspection and quantitative measurement.

For evaluation, we choose 3 subjects, and manually segment all their 5 time-points by a trained rater. The proposed segmentations are compared with manual ground truth using the Dice ratio, $DR(L,M) = 2|L \cap M|/(|L| + |M|)$, where L is the estimated segmentation label and M is the manual segmentation label. Dice ratio ranges from 0 to 1. For comparison, the initial segmentations are used, which are performed by a longitudinal segmentation method described in [4]. Briefly, the fuzzy segmentation result of the 12-month T1 image is used as a prior to guide the

segmentations of all other time-points, respectively. Unlike our method, it is a one-way prior propagation and only two time-points can be involved. Their results are referred to as "LongSeg-T1" and "LongSeg-T2". Segmentations of a representative infant are shown in Fig. 4. Visually, our results show better anatomical details and are more consistent compared with the two control methods without temporal consideration, by taking manual segmentations as reference. The quantitative results are shown in Fig. 5. The accuracy of the independent segmentation on T1 images is initially low, but is eventually higher. A reverse pattern can be observed for T2 images. The proposed method is robust and achieves the best segmentation accuracy.

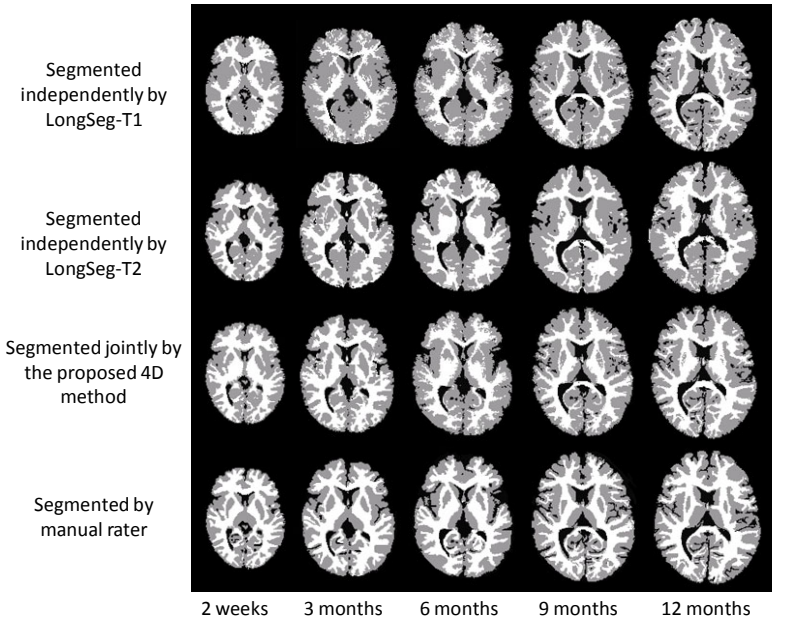


Fig. 4. Segmentation results of LongSeg-T1, LongSeg-T2, the proposed method, and a manual rater. The original T1 and T2 images are shown in Fig. 1.

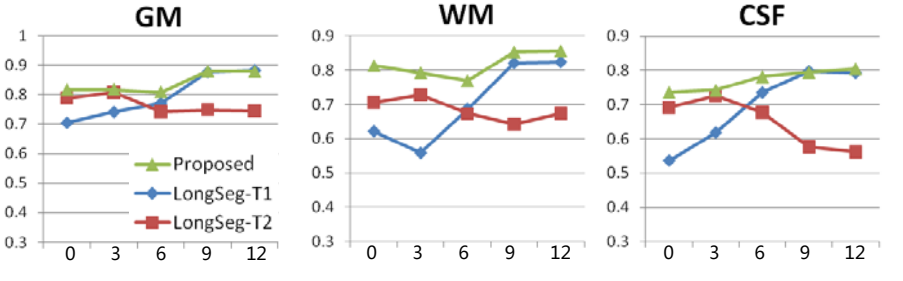


Fig. 5. Mean Dice ratios (y-axis) of 3 subjects for the LongSeg-T1, LongSeg-T2 and the proposed method, compared with the manual segmentations. X-axis is the postnatal month when image acquired.

4 Conclusion

A 4D joint registration and segmentation framework is proposed in this paper. The goal is to improve segmentation of infant images from the isointense stage, and at the same time preserve temporal consistency. Spatial-temporal constrain is introduced in the construction of prior probabilistic maps, by making optimal use of both T1 and T2 images, and by adaptively borrowing knowledge from the neighboring time-points. 4D registration and multi-channel segmentation are iterated as an EM process, assisting each other to progressively generate better results. Experimental results indicate that our method yields the highest agreement with the manual rater, and outperforms other methods in comparison.

Different with the longitudinal segmentation method in [4], the proposed method considers significantly more time-points, with the ability to take advantage of information from both temporally increasing and decreasing directions. It also incorporates a more principled structural-temporal constraint, effectively gathering more information spatially and temporally for accurate segmentation of not only neonatal images but also images at the 6 months time-point, which typically have images with very low tissue contrast. The proposed method will be further tested using a large set of longitudinal infant images currently acquired in the authors' institute, and is expected to facilitate longitudinal infant brain development research by providing an effective means for segmenting infant images from the isointense stage with temporal consistency.

References

1. Knickmeyer, R.C., Gouttard, S., Kang, C., Evans, D., Wilber, K., Smith, J.K., Hamer, R.M., Lin, W., Gerig, G., Gilmore, J.H.: A structural MRI study of human brain development from birth to 2 years. J. Neurosci. 28, 12176–12182 (2008)
2. Prastawa, M., Gilmore, J.H., Lin, W., Gerig, G.: Automatic segmentation of MR images of the developing newborn brain. Medical Image Analysis 9, 457–466 (2005)
3. Xue, H., Srinivasan, L., Jiang, S., Rutherford, M., Edwards, A.D., Rueckert, D., Hajnal, J.V.: Automatic segmentation and reconstruction of the cortex from neonatal MRI. Neuroimage 38, 461–477 (2007)
4. Shi, F., Fan, Y., Tang, S., Gilmore, J.H., Lin, W., Shen, D.: Neonatal brain image segmentation in longitudinal MRI studies. Neuroimage 49, 391–400 (2010)
5. Dietrich, R.B., Bradley, W.G., Zaragoza, E.J.t., Otto, R.J., Taira, R.K., Wilson, G.H., Kangarloo, H.: MR evaluation of early myelination patterns in normal and developmentally delayed infants. American Journal of Roentgenology 150, 889–896 (1988)
6. Shattuck, D.W., Leahy, R.M.: Automated graph-based analysis and correction of cortical volume topology. IEEE Transactions on Medical Imaging 20, 1167–1177 (2001)
7. Sled, J.G., Zijdenbos, A.P., Evans, A.C.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Transactions on Medical Imaging 17, 87–97 (1998)
8. Ashburner, J., Friston, K.J.: Unified segmentation. Neuroimage 26, 839–851 (2005)
9. Shen, D., Davatzikos, C.: Measuring temporal morphological changes robustly in brain MR images via 4-dimensional template warping. Neuroimage 21, 1508–1517 (2004)
10. Sundar, H., Litt, H., Shen, D.: Estimating myocardial motion by 4D image warping. Pattern Recognition 42, 2514–2526 (2009)

Multi-parametric Classification of Traumatic Brain Injury Patients Using Automatic Analysis of Quantitative MRI Scans

Benjamin S. Aribisala[1], Christopher J.A. Cowie[1], Jiabao He[1], Joshua Wood[1], A. David Mendelow[1,2], Patrick Mitchell[1,2], and Andrew M. Blamire[1]

[1] Institute of Cellular Medicine, Newcastle MR Centre, Newcastle University, UK
[2] Dept of Neurosurgery, Newcastle General Hospital, Newcastle upon Tyne, UK

Abstract. Traumatic brain injury (TBI) is ranked as the fourth highest cause of death in the developed world. The majority of patients sustain mild TBI, and a significant number suffer persistent neuropsychological problems. Conventional neuroimaging methods (CT, MRI) do not reveal abnormalities consistent with the cognitive symptoms. Imaging methods offering prognostic information in acutely injured patients are therefore required. Here we applied advanced quantitative MRI techniques (T_1, T_2 mapping and diffusion tensor MRI) in 24 mild TBI patients and 20 matched controls. We applied a support vector machine (SVM) to classify the quantitative MRI data. Univariate classification was ineffective due to overlap between patient and control values, however multi-parametric classification achieved sensitivity of 88% and specificity of 75%. Future work incorporating neuropsychological outcome into SVM training is expected to improve performance. These results indicate that SVM analysis of multi-parametric MRI data is a promising approach for predicting prognosis following mild TBI.

Keywords: Traumatic Head Injury, Support Vector Machine, Magnetic Resonance Imaging, Multi-parametric, TBI, relaxometry, DTI, T1, T2.

1 Introduction

Traumatic brain Injury (TBI) is a significant cause of death and disability in adults. Each year in the UK more than 112,000 people are admitted from accident and emergency departments with a primary diagnosis of TBI [1]. TBI is ranked as the fourth highest cause of death in the developed world, and the number of people sustaining head injuries increases yearly [2]. Around 90% of admissions for head injury are classed as mild, with the remainder divided equally between moderate and severe injuries [3]. Focussing on the mild injury group it is clear that a considerable number of patients have cognitive difficulties related to their injury, including problems with concentration and memory[4] which persist for months if not years. Developing non-invasive methods which provide early prognostic information is an important goal in this patient group.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 51–59, 2010.

Computed tomography (CT) is used for initial assessment of TBI patients but CT and conventional MR imaging (e.g. T_1W and T_2W) in *mild* TBI patients often does not correlate with the severity and longevity of the clinical neurological picture [5, 6]. It has been reported in small cohort studies of TBI that advanced MRI techniques such as diffusion tensor imaging (DTI) and image relaxometry do detect subtle quantitative changes in brain tissue properties [5, 7, 8], but single measurements do not have prognostic value in individual patients. In view of these previous findings we anticipate that combination of a range of quantitative MRI parameters will be more sensitive in detection of subtle neuronal damage than when using single parameters. Hence we hypothesised that multi-parametric analysis would offer a better prediction of tissue damage in TBI patients than univariate analysis. In order to test our hypothesis we applied the machine learning classification method of Support Vector Machines (SVMs).

SVMs are supervised classification algorithms which use training sets to learn the differences between groups to be classified. The wide ranging application of SVM to biological problems is due its high accuracy during classification, the ability to deal with multi-dimensional and large datasets and the high flexibility in modelling of data from various sources [9]. Current applications of SVM include classification of patients and control groups in diabetes mellitus [10] and Alzheimer disease [11].

2 Materials and Methods

2.1 Subjects

A total of 44 subjects were recruited for this study. This comprised 24 mild TBI patients (Glasgow Coma Score, GCS 14-15, mean age 38±15 yrs) and 20 healthy adults (mean age 41±16 yrs) with no clinical evidence of neurological diseases or prior history of TBI. Scanning in the patient group was performed within 10 days of injury (mean 4.9, range 1 -10 days).

2.2 MR Imaging and Protocol

Human brains are anatomically divided into gray matter (neuronal cell body), white matter (axon fibres) and cerebral spinal fluid (CSF), and MR signal characteristics arising from each compartment are distinctly different. T_1 weighted anatomical images (T_1W) are the standard clinical high resolution anatomical images commonly used for visual assessment due to the high contrast delineating GM, WM and CSF from each other. T_1W imaging is useful for image registration and definition of regions of interests. However, the information provided by T_1W is qualitative because it depends on both tissue properties and scanner performance factors.

The magnetic resonance relaxation times (longitudinal relaxation time, T_1 and transverse relaxation time, T_2) are indicators of the local tissue micro-environment experienced by water molecules from which the MR signal originates and which are altered by disease or injury. Measurement of T_1 and T_2 provides scanner independent quantitative information reflecting fundamental tissue properties.

Diffusion tensor imaging (DTI) probes the tissue microstructure through Brownian motion of water molecules, estimating the mean diffusivity (MD), an indicator of the

level of diffusion restriction and tissue boundaries. MD provides quantitative information about the tissue microstructure and it is well known that the MD values are also sensitive to disease processes.

We anticipate that combination of these three quantitative images (T_1, T_2 and MD) will lead to improved classification of patient or disease groups from healthy controls. In view of this we acquired the sets of images with details given below. All images were acquired on a 3.0T whole body Philips Achieva MR System (Philips Medical Systems, Best, NL) using an 8-channel SENSE head coil. The protocol was approved by the local ethical committee and all subjects provided written consent prior to imaging. The following scans were acquired in each subject.

T_1W Imaging: High resolution 3D T_1 weighted anatomical scan (Magnetization Prepared Rapid Gradient sequence, TR/TE=8.1/4.6ms, matrix size 150×240 with 240 contiguous slices, 1mm slice thickness, in-plane resolution of 1mm).

T_1 Mapping: A fast quantitative T_1 measurement using a custom inversion recovery prepared Echo Planar Imaging sequence (TR/TE=15s/24ms, TIR=0.25 to 2.5s in uniform 12 steps, matrix 128×128, 72 axial slices, isotropic 2mm resolution).

T_2 Mapping: A fast quantitative multi-echo T_2 measurement based on a Gradient and Spin Echo Imaging sequence (TR=4.7s, 8 spin echoes at 20ms spacing, EPI factor 5, matrix 128x128, 72 slices, isotropic 2mm resolution).

Diffusion Tensor Imaging: DTI using a spin-echo, echo-planar sequence (SENSE factor 2, TE/TR=71/2524ms, matrix 128x128, 24 slices, 6 mm thickness and 2mm in-plane resolution, 16 diffusions directions, b values 0 and 1000 s mm^{-2}).

B_o Field-map: MR images are subjected to geometry distortions resulting from magnetic field variations across the subject. A B_o field-map (dual echo 3D Gradient Echo sequence TR=27ms, TE=2.6/6.1ms, matrix 128×128×72, 2mm resolution) was collected to correct the spatial distortion in Echo Planar Imaging based images.

2.3 Image Analysis

Patients with visible lesions were removed before analysis. We therefore analysed only normal appearing tissues. We expected to find micro structural damage in areas known to be affected by diffuse axonal injury. In mild TBI, damage can be slight and whole brain analysis is not sensitive enough in these cases. We therefore increased the likelihood of detection by subdividing the brain into the specific regions of interests.

We applied an automatic image analysis method whereby the whole brain is automatically divided into 16 regions of interest (ROI) for each tissue type [12]. These regions are pairs of right and left *inferior frontal lobe, superior frontal lobe, temporal lobe, temporal-occipital lobe, occipital lobe, temporal-parietal lobe, parietal lobe and the cerebellum*. This analysis method operates in patient space which was demonstrated to significantly reduce partial volume errors compared to the same analysis performed in standard space [13]. In brief, the method uses a set of standard space brain ROIs parcellating the entire brain, which are transformed into patient space based on a multi-step registration using the patient's high resolution T_1 weighted anatomical scan. Next, the same anatomical scan is segmented into white matter, grey matter and CSF masks [14] and combined with the brain region template

to generate tissue specific anatomical ROIs which are applied to the quantitative images under analysis. Multi-spectral analysis using k-means clustering is applied to the regional quantitative data for removal of partial volume errors in order to improve ROI definitions. The algorithm was implemented in MATLAB R2009b (The MathWorks Inc., Natick, MA 01760-2098, USA) running on a Linux platform using in-house developed routines but incorporated existing processing methods from the FSL [15] package when appropriate. All segmentation steps were performed using FSL Segmentation Tool (FAST, [14]).

T_1 and T_2 Mapping: Quantitative T_1 maps (qT_1) were calculated on a pixel by pixel basis by fitting the acquired data to T_1 inversion recovery curve using the standard 3 parameter fit (M_o, flip angle and T_1) while quantitative T_2 maps (qT_2) were calculated using a 2 parameter (M_o and T_2) monoexponential fit to the acquired data.

Diffusion Tensor Imaging: DTI data were preprocessed with FDT (FMRIB's Diffusion Toolbox) [16]. Head movement and eddy currents were corrected using 3D rigid body registration to a reference volume. Raw DTI data were brain-extracted using FSL BET tool, and mean diffusivity (MD) images were created by fitting a tensor model to the raw diffusion data using FDT.

The algorithm was then used to automatically determine regional grey and white matter qT_1, qT_2, and MD in each of the 16 target ROIs. Finally, the regional mean values for both grey and white matter were computed in each ROI and used for SVM classification.

2.4 Support Vector Classification of TBI Data

SVM was used to classify the regional mean values computed from qT_1, qT_2 and MD. Each subject's data was divided into the 2 tissue classes of interest (grey matter and white matter) with each comprising of 16 x 3 matrices, representing the 16 ROIs for each of the 3 quantitative MRI parameters. These matrices were used as input vectors for SVM. Each of the two groups (mild TBI and control) was divided into 2 mutually exclusive subsets, the training set and the validation set. Selection was done using the holdout cross validation method; this method randomly divides a given dataset into 2 equal groups. Training and classification were evaluated in each ROI for both white matter (WM) and grey matter (GM) using combinations of qT_1 and qT_2, qT_1 and MD, qT_2 and MD and qT_1, qT_2 and MD. We compared a number of kernel functions using sensitivity and specificity analysis, only the radial basis function gave a desirable result. In view of this finding (no data presented) our implementation used radial basis function.

3 Results and Discussions

Table 1 shows the mean and standard deviation for qT_1 qT_2 and MD in selected regions of interests. This result shows that values of T_1WM, T_1GM, T_2GM and MDWM were elevated in almost all ROIs in the patients compared with the control group. This result agrees with previous work showing increase in mean diffusivity [8]. Statistically significant group differences ($p<0.05$) were found in White matter T_2

(left parietal), White matter MD (left frontal superior, left temporal parietal and right frontal superior), Grey matter T_1 (right temporal, right occipital and left temporal parietal) and Grey matter T_2 (right temporal and right occipital).

Figure 1 shows selections from typical control and patient datasets. Figure 2 shows a representative scatter plot and SVM results. The scatter plots show that there is significant overlap between the 2 groups along each axis but that combination of axes reveals some intra-group relationships. The SVM results on the right hand side of each plots show the separation between groups. We used sensitivity (True positive) and specificity (True negative) to measure the performance of SVM.

Table 1. Quantitative T_1 T_2 and MD (mean and standard deviation) values in 20 control subjects and 20 patient subjects (* $p<0.05$) for Selected Regions of Interests

Controls Group- (T_1 and T_2 in ms, MD×10^{-6} mm^2s^{-1})						
ROI	T_1WM	T_2WM	MDWM	T_1GM	T_2GM	MDGM
Right Fron.	813 ±34	84 ±5	680 ±31	1164 ±51	90 ±5	822 ±39
Left Fron.	811 ±33	84 ±5	673 ±29	1158 ±46	90 ±5	824 ±48
Right Temp.	805 ±29	78 ±2	830 ±26	1216 ±36	90 ±3	770 ±125
Right Occi.	804 ±40	82 ±2	775 ±21	1038 ±62	81 ±3	846 ±38
Right Temp.	794 ±30	81 ±2	781 ±25	1251 ±22	90 ±3	926 ±47
Left Temp.	806 ±46	80 ±2	764 ±24	1250 ±23	91 ±3	919 ±37
Left Parietal	819 ±34	84 ±2	749 ±37	1182 ±40	86 ±4	873 ±43
Patients Group - (T_1 and T_2 in ms, MD×10^{-6}						
ROI	T_1WM	T_2WM	MDWM	T_1GM	T_2GM	MDGM
Right Fron.	819 ±46	83 ±4	**708 ±24***	1181 ±109	91 ±5	814 ±57
Left Fron.	817 ±45	83 ±4	**701 ±21***	1181 ±84	91 ±4	824 ±72
Right Temp.	818 ±36	78 ±2	842 ±41	**1279 ±116***	**94 ±6***	808 ±54
Right Occi.	811 ±36	82 ±4	779 ±36	**1140 ±163***	**86 ±6***	836 ±38
Right Temp.	798 ±38	80 ±2	793 ±26	**1275 ±58***	91 ±4	918 ±62
Left Temp.	806 ±44	80 ±3	**781 ±23***	1271 ±35	91 ±2	927 ±58
Left Parietal	818 ±45	**82 ±2***	763 ±25	1207 ±58	87 ±5	875 ±56

Table 2. Average Sensitivity and Specificity as a measure of performance for SVM

Combinations	Control Vs Mild TBI			
	White Matter		Grey Matter	
	Sensitivity	Specificity	Sensitivity	Specificity
T_1 and T_2	82	70	80	75
T_1 and DTI	82	70	87	79
T_2 and DTI	81	73	88	75
T_1,T_2 and DTI	83	68	85	81

Table 2 shows the average sensitivity (and specificity) for both white matter and grey matter averaged across all the 16 regions of interests. For white matter the average sensitivity ranged from 81% to 83% (specificity from 68 to 73) and from 80 to 88% (specificity from 75 to 81) for grey matter. The table also shows that for white matter regions, combining 3 parameters gives a higher sensitivity as compared with the combination of 2 parameters. For grey matter regions the highest sensitivity is achieved

when T_2 and MD are combined. It should be stressed that the analysis considers only normal appearing brain tissue and hence these results suggest that multi parametric analysis using SVM offers a promising tool for detecting tissue damage in mild TBI.

Epidemiologically, only approximately half of mild TBI patients manifest ongoing neuropsychological problems related to their injury [4]. In view of this, approximately 50% of the TBI population is expected to be indistinguishable from normal controls and this could cause misclassification. We believe that this may be a significant contributing factor to the low specificity of our analysis. Our future work will include follow up studies in order to identify the mild TBI patients who have fully recovered without any neuropsychological symptoms which will help us to redefine the groups which could lead to improved specificity. Our future work will also attempt to increase the sample size for both the control and the patient group in order to increase the statistical power of the classification method.

SVM has been used in a few studies to investigate patients with TBI [5, 17, 18] and it is important to place our findings in context with existing work. Mourao-Miranda et al. [17] performed a study based on functional MR to investigate brain activity and did not evaluate structural changes while Davies et. al. [18] used only categorical data such as age, gender, injury severity and GCS scores. In these studies, only Meyer et al. [5] used MRI parameters in a fashion similar to our study.

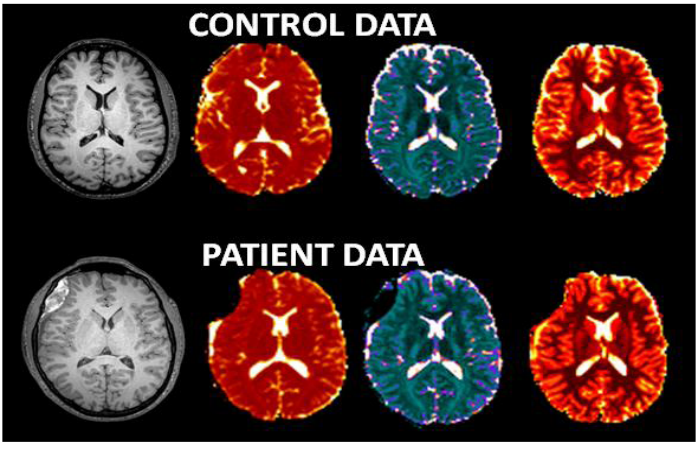


Fig. 1. Typical quantitative images representing the control and patient populations. Columns 1, 2, 3 and 4 are the T_1W, MD, qT_2 and qT_1 images respectively

Meyer et al. [5] used SVM to evaluate multiple data in mild TBI patients. In that study, Fractional Anisotropy (FA, another metric derived from diffusion tensor MRI) was combined with clinical measures (e.g. attention, memory, emotional complaints etc). According to Meyer et al. the use of clinical measures did not improve classification which implies that, technically speaking, their classification is based only on FA. Additionally, they evaluated 3 specific regions of interests namely the

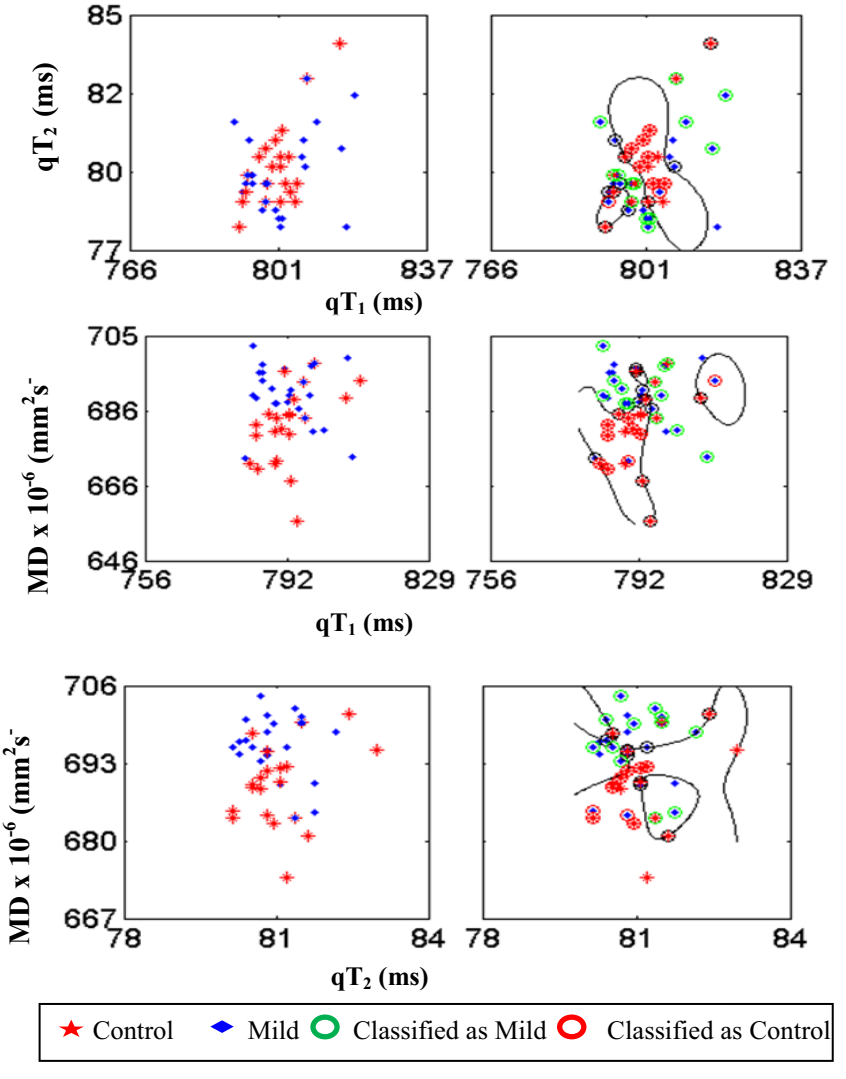


Fig. 2. The scatter plot and SVM results in the Frontal Superior region of white matter. First row qT_1 against qT_2, Second row qT_1 against MD and third row qT_2 against MD.

Corpus Callosum and left and right hemisphere in the white matter tissue class. Although we did not include FA in our current analysis, our data suggest that single MR parameters do not show sufficient separation to be prognostic. In our study we therefore combined 3 quantitative MR parameters (qT_1, qT_2 and MD) and our analysis covered all the brain regions within both white and grey matter tissue classes. Our results show that there are regional changes not only in white matter but also in grey matter regions. Extending the dimensionality of the data gave sensitivity (and

specificity) of 88% (and 75%) which is an improvement on previously reported findings (81% and 63% respectively [5]).

To the best of our knowledge we are the first to apply the robust SVM method for classification of mild traumatic brain injury patients from control subjects based on automatic analysis of multiple quantitative magnetic resonance parameters.

4 Conclusions

We have shown that a multi-parametric analysis of quantitative MRI data can be used to separate mild TBI patients from the control group. Our results show that SVM can detect changes in normal appearing tissues in some patients suffering mild TBI as compared with the control group. These changes may represent damage to neuronal tissue and further work is needed to determine whether this is responsible for the cognitive and affective symptoms commonly seen following mild head injury, which include memory loss, inability to concentrate, irritability and depression.

References

[1] NICE, Clinical Guideline 56. Head Injury: Full Guideline. London: National Institute for Health and Clinical Excellence (2007)
[2] Reilly, P., Bullock, R.: Head injury pathophysiology and management. Hodder Arnold (2005)
[3] Kay, A., Teasdale, G.: Head injury in the United Kingdom. World Journal of Surgery 25(9), 1210–1220 (2001)
[4] Nolin, P., Heroux, L.: Relations Among Sociodemographic, Neurologic, Clinical, and Neuropsychologic Variables, and Vocational Status Following Mild Traumatic Brain Injury: A Follow-up Study. Journal of Head Trauma Rehabilitation 21(6), 514–526 (2006)
[5] Mayer, A.R., Ling, J., Mannell, M.V., et al.: A prospective diffusion tensor imaging study in mild traumatic brain injury. Neurology 74, 643–650 (2010)
[6] Lee, H., Wintermark, M., Gean, A.D., et al.: Focal Lesions in Acute Mild Traumatic Brain Injury and Neurocognitive Outcome: CT versus 3T MRI. Journal of Neurotrauma 25(9), 1049 (2008)
[7] Lee, H., Wintermark, M., Gean, A., et al.: Focal Lesions in Acute Mild Traumatic Brain Injury and Neurocognitive Outcome: CT versus 3T MRI. Journal of Neurotrauma 25(9), 1049 (2008)
[8] Goetz, P., Blamire, A., Rajagopalan, B., et al.: Increase in apparent diffusion coefficient in normal appearing white matter following human traumatic brain injury correlates with injury severity. Journal of Neurotrauma 21(6), 645–654 (2004)
[9] Schölkopf, B., Tsuda, K., Vert, J.: Kernel methods in computational biology. MIT Press, Cambridge (2004)
[10] Barakat, N., Bradley, A., Barakat, M.: Intelligible Support Vector Machines for Diagnosis of Diabetes Mellitus. IEEE Trans. Inf. Technol. Biomed. (2010)
[11] Klöppel, S., Stonnington, C., Chu, C., et al.: Automatic classification of MR scans in Alzheimer's disease. Brain 131(3), 681–689 (2008)
[12] Aribisala, B.S., He, J., Thelwall, P.E., Hollingsworth, K.G., Blamire, A.M.: On Automatic Regional Analysis of Quantitative Relaxation Times Mapping in the Brain. In: International Society for Magnetic Resonance in Medicine, vol. 16 (2008)

[13] Aribisala, B.S., Blamire, A.: Comparison of Analysis of Brain Relaxation Times in Standard Space with Analysis in Individuals' Real Space. In: Conference, International Society for Magnetic Resonance in Medicine, Honolulu, USA (2009)
[14] Zhang, Y., Brady, M., Smith, S.: Segmentation of brain MR images through a HMRFmodel and the expectation-maximization algorithm. IEEE Transactions on Medical Imaging 20(1), 45–57 (2001)
[15] FMRIB Software Library, A comprehensive library of analysis tools for FMRI, MRI and DTI brain imaging data. Written by members of the Analysis Group, FMRIB, Oxford, UK, Oxford University (2007)
[16] Behrens, T.E.J., Woolrich, M.W., Jenkinson, M., et al.: Characterization and Propegation of Uncertainty in Diffusion Weighted MR images. Magn. Reson. Med. 50(5), 1077–1088 (2003)
[17] Mourao-Miranda, J., Bokde, A.L.W., Born, C., et al.: Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data. Neuroimage 28(4), 980–995 (2005)
[18] Davis, D., Peay, J., Good, B., et al.: Air medical response to traumatic brain injury: A computer learning algorithm analysis. Journal of Traumatic-Injury Infection and Critical Care 64(4), 889–897 (2008)

Deformable Vessel-Based Registration Using Landmark-Guided Coherent Point Drift

Yipeng Hu, Erik-Jan Rijkhorst, Richard Manber, David Hawkes, and Dean Barratt

Centre for Medical Image Computing, University College London, London, UK

Abstract. Automatic, non-rigid registration of blood vessels (and other tubular structures) within a timescale suitable for use in image-guided surgical applications remains a significant challenge. We describe a novel approach to this problem in which an extension to the coherent point drift (CPD) algorithm is developed to enable landmarks, such as vessel bifurcations, to improve the registration accuracy and speed of execution. The new method – referred to as landmark-guided CPD (LGCPD) – is validated using vessels extracted from brain MRA and liver MR images, and is shown to be robust to missing vessel segments and noise, commonly encountered in realworld applications.

1 Introduction

Blood vessels, and other tubular structures, such as ducts, appear commonly in medical images. They can therefore provide very useful features for driving image registration schemes, particularly in image-guided surgical applications where knowledge of the location of blood vessels relative to surgical instruments is extremely important. Typical applications include neurosurgery, liver resection, liver transplant surgery, and vascular interventions where image guidance and registration technology is becoming increasingly relied upon to implement novel minimally-invasive techniques. Since vessels are distributed throughout organs, when they can be imaged, they provide a useful feature for capturing non-rigid deformation fields. However, modality-specific artefacts, organ deformation, missing vessel segments, and the different grey-level intensity characteristics of vessels imaged using different modalities mean that vessel-based registration in general remains a challenging task. This is particularly the case in the interventional setting where robustness, accuracy and speed are key criteria for clinical acceptance.

Vessel-based registration techniques have been reported for a number of research groups, principally those involving the liver or brain [1-4]. Many of the existing algorithms for vessel-based algorithms are feature-based approaches, which require prior segmentation of vessels and vessel centrelines [2;3]. Such methods have the advantage that the registration can often be performed very rapidly, but there is the additional overhead of the segmentation task, which may require manual interaction and can be subject to error. Intensity-based approaches, on the other hand, such as those described in [4-6] do not require prior image segmentation, are relatively computationally intensive, making them difficult to apply in many realtime

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 60–69, 2010.

applications. Aylward et al. [1], Penney et al. [7] and Lange et al. [3] all combine feature- and image-intensity-based registration approaches.

Realtime, non-rigid vessel-based registration with little or no user interaction remains a significant challenge; Lange et al. [2;3] report a computation time of 10min for non-rigid registration, excluding the time taken to segment vessel images and manually identify bifurcations, whilst Reinersten et al. [4] report that registrations were completed within 1 minute. Both of these studies adopt thin-plate splines (TPS) to approximate the non-rigid component of the registration transformation that matches vessel centrelines. In this paper, we propose an alternative approach for fast, non-rigid vessel registration, based on the recently introduced coherent point drift (CPD) method [8]. In the CPD framework, it is assumed that the points in one dataset are generated by a mixture model centered on the points in the transformed dataset. The expectation maximization (EM) algorithm can then be used to solve the maximum likelihood estimation (MLE) problem, where the point correspondences and transformation are updated in the E and M steps, respectively. A solution from the calculus of variations is derived in [8] that allows efficient estimates of the regularised non-rigid transformation, where a regularisation term comes from a prior.

CPD has the advantage of flexibility since most existing general-purpose rigid and non-rigid point-based registration schemes can be formulated using the framework. In practice, however, CPD assumes no specific point correspondence other than the one derived from the Euclidean distances between points. As a result, it is insufficiently constrained in situations where, for example, physical constraints prevent certain types of deformation. In such cases, the likelihood function may not have a well-defined maximum. Furthermore, because the assumption of a Gaussian Mixture Model (GMM) may be inadequate or incorrect with imperfect data, leading to false global maxima in the likelihood function, the algorithm can be sensitive to the presence of outliers and/or missing data.

To address these limitations, we extend the original CPD algorithm so that explicitly defined, corresponding pairs of anatomical landmark points can be introduced, leading to improved behaviour and registration accuracy compared with the standard algorithm. Automatic registration of blood vessels using the *landmark-guided* CPD (LGCPD) algorithm is demonstrated using images of cerebral and liver vessels.

2 Methods

2.1 Vessel Segmentation

In this study, vessels to be registered were represented in two forms: a set of vessel surface points and a set of points that define the vessel centreline. These data were extracted from vessel images as follows: (1) the vessel image was first enhanced using an ITK implementation of a multi-scale vessel enhancing filter [9]. The outputs of this filter are a ‘vesselness’ image in which the intensity of each voxel provides a measure of the likelihood of a locally vessel-like feature being present in that voxel, and a 3D vector field, where the 3D local vessel direction vector is defined at each voxel [10]; (2) the vesselness image is converted to a binary volume, B, by applying a

user-defined threshold; (3) the surface pointset was sampled from the iso-surface computed from B; (4) the centrelines were extracted from B using the 3D thinning, described in [11].

2.2 Coherent Point Drift Framework

The CPD method provides a framework for rigid and non-rigid alignment of two pointsets, where a target pointset is represented by 'data' vectors, $\mathbf{x}_n$ ($n=1,2,\ldots,N$), and a source pointset is represented by 'model' vectors, $\mathbf{y}_m$ ($m=1,2,\ldots,M$). Both vectors contain D-dimensional coordinates ($D=3$ here). Each transformed model vector $T(\mathbf{y}_m, \boldsymbol{\theta})$ is considered to contain the centroids of a set of Gaussians, where T is the spatial transformation of $\mathbf{y}_m$ with transformation parameters $\boldsymbol{\theta}$. This pointset forms a mixture of M Gaussians with a weighted uniform distribution $p(\mathbf{x}|m=M+1)=1/N$ (with scalar weight ω, $0 \le \omega \le 1$) to explicitly account for outliers. The probability density function of the random vector $\mathbf{x}$ is then given by:

$$p(\mathbf{x}) = \sum_{m=1}^{M+1} P(m)\, p(\mathbf{x} \mid m) = \omega \frac{1}{N} + (1-\omega) \sum_{m=1}^{M} \frac{1}{M} G(\mathbf{x} \mid T(\mathbf{y}_m, \boldsymbol{\theta}), \sigma^2), \tag{1}$$

where the isotropic Gaussian distribution with parameter σ and mean vector $\boldsymbol{\mu}$ is defined as $G(\mathbf{x}|\boldsymbol{\mu},\sigma^2)=(2\pi\sigma^2)^{-D/2}\exp(-||\mathbf{x}-\boldsymbol{\mu}||^2/2\pi\sigma^2)$ Given the pointset $\mathbf{x}_n$, representing the observed data, the following negative log-likelihood function for the parameters, $\boldsymbol{\theta}$ and σ, can be used as an objective function and measure of the goodness of alignment:

$$E = -\sum_{n=1}^{N} \log p(\mathbf{x}_n) \tag{2}$$

Registration, which requires estimates of the transformation parameters, becomes a MLE problem and may be efficiently solved by the EM algorithm. The rigid or affine transformations are parameterised by a D-dimensional translation plus a D-dimensional rotation, or a $D \times D$ unconstrained matrix, respectively. Closed-form solutions for these cases are derived in [8].

As explained in [8], an appropriate prior for a non-rigid transformation is $P(v)=\exp(-(\lambda/2)||Pv||^2)$, where v is a displacement function, which defines a non-rigid deformation field, and P is a linear operator that acts in this case as a high-pass filter. Using this prior, the objective function becomes $E-\log(P(v))=E+(\lambda/2)||Pv||^2$. The EM algorithm can again be employed to solve this maximum a posterior problem so that 1) the E steps remain the same as in MLE; and 2) the negative complete log-likelihood function – the objective function to minimise in M steps – takes the following form:

$$Q(v,\sigma^2) = \frac{1}{2\sigma^2} \sum_{n=1}^{N} \sum_{m=1}^{M} P_{mn} \left\| \mathbf{x}_n - (\mathbf{y}_m + v(\mathbf{y}_m)) \right\|^2 + \frac{N\mathbf{p}D}{2} \log \sigma^2 + \frac{\lambda}{2} \left\| Pv \right\|^2, \tag{3}$$

where the terms independent of σ and v are ignored and $N\mathbf{p} = \sum_{n=1}^{N} \sum_{m=1}^{M} P_{mn}$. The posterior probabilities, P_{mn} (also known as the responsibilities or membership probabilities) are computed in E steps, and are given by:

$$P_{mn} = P^{old}(m \mid \mathbf{x}_n) = \frac{\exp(-\frac{1}{2\sigma^2}\left\|\mathbf{x}_n - T(\mathbf{y}_m, \boldsymbol{\theta}^{old})\right\|^2)}{\sum_{k=1}^{M}\exp(-\frac{1}{2\sigma^2}\left\|\mathbf{x}_n - T(\mathbf{y}_k, \boldsymbol{\theta}^{old})\right\|^2) + (2\pi\sigma^2)^{D/2}\frac{\omega}{1-\omega}\frac{M}{N}}, \tag{4}$$

where $\boldsymbol{\theta}^{old}$ are the parameters estimated in a previous iteration.

According to the calculus of variations approach and motion coherence theory [8], CPD chooses a Gaussian kernel with one parameter β, namely $k(\mathbf{y}_i, \mathbf{y}_j) = \exp(-\|\mathbf{y}_i - \mathbf{y}_j\|^2/2\beta^2)$. Thus, the optimal function form of v that minimises eq.(3) becomes a linear combination of the Gaussian kernel and M sets of D-dimensional coefficients. Therefore, the coefficients can be estimated by solving a linear least-squares problem and σ^2 may be estimated by equating the corresponding derivative to zero. In [8], it is also suggested that a good minimum generally can be achieved with a single iteration of these two evaluations. A more detailed description of the solution and the implementation are omitted here for brevity, but can be found in [8].

2.3 Landmark-Guided Coherent Point Drift

In this section, we consider the situation where L additional landmark pairs are available to constrain the registration. In practice, such landmarks are treated differently from other vessel points. If $\mathbf{x}_l^*$ and $\mathbf{y}_l^*$ contain the co-ordinates of corresponding landmarks for $l=1,2,\ldots,L$, we can in turn define a mathematically convenient prior over $\boldsymbol{\theta}$ as the following joint probability of L Gaussians:

$$P(\boldsymbol{\theta} \mid \sigma^{*2}) = \prod_{l}^{L} G(\mathbf{x}_l^* \mid T(\mathbf{y}_l^*, \boldsymbol{\theta}), \sigma^{*2}) \tag{5}$$

where σ^{*2} is the new hyper-parameter, which controls the influence of the prior. Furthermore, in the non-rigid case, by assuming the independence between the known correspondence and the prior on the displacement field $P(v)$, the new "joint" prior may be given by:

$$P(v \mid \sigma^{*2}) = \prod_{l}^{L} G(\mathbf{x}_l^* \mid (\mathbf{y}_l^* + v(\mathbf{y}_l^*)), \sigma^{*2}) \cdot \exp(-\frac{\lambda}{2}\left\|Pv\right\|^2) \tag{6}$$

Therefore, the new objective function to minimise in M steps becomes:

$$Q^*(v, \sigma^2) = Q(v, \sigma^2) - \frac{\lambda}{2}\left\|Pv\right\|^2 - \log(P(v \mid \sigma^{*2})) \tag{7}$$

where Q is defined as in eq.(3). To obtain the transformation minimising eq.(7), the same approach described in Section 2.2 may be used. Taking the functional derivative of Q^*, so the independent terms may be omitted, gives:

$$\begin{aligned}\frac{\partial Q^*}{\partial v} &= \frac{\partial}{\partial v}[\frac{1}{2\sigma^2}\sum_{n=1}^{N}\sum_{m=1}^{M}P_{mn}\left\|\mathbf{x}_n - (\mathbf{y}_m + v(\mathbf{y}_m))\right\|^2 + \frac{N\mathbf{p}D}{2}\log\sigma^2 + \frac{\lambda}{2}\left\|Pv\right\|^2 \\ &\quad + \frac{1}{2\sigma^2}\sum_{l=1}^{L}\frac{\sigma^2}{\sigma^{*2}}\left\|\mathbf{x}_l^* - (\mathbf{y}_l^* + v(\mathbf{y}_l^*))\right\|^2] \\ &= \frac{\partial}{\partial v}[\frac{1}{2\sigma^2}\sum_{j=1}^{N+L}\sum_{i=1}^{M+L}P_{ij}^*\left\|\mathbf{x}_j - (\mathbf{y}_i + v(\mathbf{y}_i))\right\|^2 + \frac{N\mathbf{p}D}{2}\log\sigma^2 + \frac{\lambda}{2}\left\|Pv\right\|^2]\end{aligned} \tag{8}$$

where, $P_{ij}^{*} = \begin{cases} P_{mn}, & (i = m, m \in [1, M]) \,\&\, (j = n, n \in [1, N]) \\ \sigma^2 / \sigma^{*2}, & (i = M + l) \,\&\, (j = N + l), l \in [1, L] \\ 0, & otherwise \end{cases}$

i.e. $[P_{ij}^{*}]_{(M+L)\times(N+L)} = \begin{bmatrix} [P_{mn}]_{M\times N} & \mathbf{0} \\ \mathbf{0} & \dfrac{\sigma^2}{\sigma^{*2}} \cdot \mathbf{I}_L \end{bmatrix}$,

$[\mathbf{x}_j]_{D\times(N+L)} = [[\mathbf{x}_n]_{D\times N}, [\mathbf{x}_l^{*}]_{D\times L}]$ and $[\mathbf{y}_i]_{D\times(M+L)} = [[\mathbf{y}_m]_{D\times M}, [\mathbf{y}_l^{*}]_{D\times L}]$. Eq.(4) can again be used to compute P_{mn} in E steps. As shown in the derivation of eq.(8), the new objective function has the same form of functional derivative as $\partial Q/\partial v$, so the same variational procedure to minimise Q given above may be used to minimise Q^*, with matrices $[P_{mn}]$, $[\mathbf{x}_n]$ and $[\mathbf{y}_m]$ being replaced by $[P_{ij}^*]$, $[\mathbf{x}_j]$ and $[\mathbf{y}_i]$, respectively. The same procedure also applies for rigid and affine cases, where setting the partial derivatives of the objective function in M steps to zeros leads to direct solutions for $\boldsymbol{\theta}$ and σ^2, subject to the same matrix replacements described as above. To update σ^2, equating the corresponding derivative to zero gives the estimator:

$$\hat{\sigma}^2 = \frac{1}{N\mathbf{p}D} \sum_{n=1}^{N} \sum_{m=1}^{M} \left\| x_n - (y_m + v(y_m)) \right\|^2 \tag{9}$$

The original CPD has three parameters, ω, λ and β, controlling the influence of outliers, the amount of smoothness regularisation and strength of the points' mutual interaction, respectively. The additional parameter σ^* reflects the localisation error of the corresponding landmarks. The smaller the value is set for σ^{*2}, the stronger the constraints on the corresponding landmarks. In practice, this parameter can be adjusted depending on the spatial accuracy with which landmarks are defined.

2.4 Validation

In order to evaluate the performance of the LGCPD algorithm for vessel-based registration, experiments were performed using vessels extracted from one patient brain MR image, and two liver MR images, obtained on the same volunteer. A significant issue in validating registration algorithms is establishing an accurate gold standard transformation between images that is independent of the method under evaluation. In the first set of experiments, a finite element (FE) model was used to apply a physical deformation to a volume containing cerebral vessels, segmented from a brain MR images. This volume was used as the target volume, to which the original source volume was registered. This approach has the advantage that the voxel displacement field is known and can be compared directly with displacements estimated by the vessel-based registration. In a second experiment, an MR image of a liver was deformed using a displacement field generated by registering two MR images obtained at different phases of the breathing cycle. This displacement field was then applied to one of the MR images to generate a target image. Again, the displacement field is known, and therefore provides a "perfect" gold standard for validation.

2.4.1 Brain Vessel Deformation Using a Biomechanical Simulation

Using the method described in Section 2.1, surface and centreline points from a section of a cerebral blood vessel were extracted from a MR angiography (MRA) image of a consented patient who underwent neurosurgery at the National Hospital for Neurology and Neurosurgery, London. The voxel size of the MR volume 0.49x0.49x0.7mm^3, and the image was acquired using a 3D TOF (MOTSA) sequence on a GE Signa Excite 1.5T scanner and an 8-channel SENSE phased-array head coil.

A tetrahedral mesh of the blood vessels was created from the binary image obtained by thresholding the vesselness image resulting from filtering the original MR image. This binary image was then labelled and imported to the MATLAB meshing interface tool, *iso2mesh* [12], which automatically constructs 3D solid tetrahedral meshes. As shown in Fig.1, a spherical region near to the vessel was included in the FE mesh to represent a region of interest (ROI), such as a tumour. The tissue region between the vessel and ROI was assumed to be homogeneous.

An anisotropic, nonlinear finite element (FE) solver [13] was used to simulate a physically plausible deformation of the vessel and tumour model. Displacement loads with a magnitude of 15 mm, in x, y and z directions, were applied to nodes within regions near the ends of the two vessel branches, while the other side of the mesh, near hepatic hilum, was left fixed. The tumour and surrounding tissue were assumed to be isotropic elastic materials described by a neo-Hookean model. The vessel wall was assumed to be a transversely isotropic, nonlinear material [13]. At each point along the vessel, transverse plane was defined as the plane orthogonal to the local vessel direction vector output from the vesselness filter. For the purposes of this study, all the isotropic materials were assumed to have a shear modulus of, 33.6kPa and 0.336kPa, and a bulk modulus of, 1.67GPa and 16.7kPa, for the ROI and surrounding tissue, respectively. The transverse isotropy for the vessel results in one extra stiffness parameter [13] which was set to 16.7kPa. Mesh node displacements were computed using the non-linear FE solver, described in [13] and implemented to run on a graphics processing unit (GPU).

The surface and centreline points defining the cerebral vessel were transformed into the deformed space by interpolating the FE nodal displacement field. The resulting points were considered as target points in the registrations. An additional rigid transformation was also applied before registration to test the algorithm.

2.4.2 Liver Vessel Deformation

To obtain a gold standard displacement field due to breathing motion from MR scans where the liver vasculature was clearly visible, two breath-hold MR volumes were registered using a fluid registration method [14]. This registration has been previously evaluated using real-time MR data and found to accurately model the respiratory liver motion [15]. The images were acquired on a healthy volunteer using a 1.5 T cylindrical bore Philips Achieva MR scanner at Guy's Hospital, London. The breath-hold scans were acquired at maximum exhale and mid-cycle with a field-of-view of 400x400x270 mm^3 covering the whole abdomen. The voxel size was 1.4x1.4x1.7 mm^3.

For the liver vessels, vessel surface and centreline points were extracted as described above. The right hepatic vein was identified by manually defining one point within and finding all connected voxels from the binary representations of both the source and target images.

2.4.3 Registration Experiments

To investigate the effects of missing and noisy data on the vessel-based registration, the target (deformed) cerebral vessel pointsets were modified in the following ways: First, a 7mm segment from one branch was removed at a random position. Second, random anisotropic Gaussian noise with a variance of 1mm along the vessel and 9mm orthogonal to the vessel was added to the location of each target point. Third, the first two modifications were applied simultaneously. One hundred registrations were performed for each of the three cases using the original CPD method (without landmarks) and then the LGCPD method with three pairs of manually defined landmarks: one at the bifurcation and the one at each end of the two branches. The parameter values $\sigma^*=0.01$, $\omega=1$, $\beta=1$ and $\omega=0.1$ were used in all the experiments.

For the MR liver data, registrations were performed using the same parameter values, apart from a larger smoothness parameter of $\beta=3$ due to the relatively noisy data. In LGCPD experiment, two pairs of landmarks were identified manually at two bifurcations, as shown in Fig.2.

To quantify the registration accuracy, a set of points were defined in the source images. For the cerebral vessel registration, the nodes inside the tumour and vessel surface points were used, whereas for the liver vessels the 3D volume surrounding the vessels was uniformly resampled. In each case, the displacement field produced by the CPD/LGCPD registration transformation was used to propagate the point co-ordinates into the target space. The target registration error (TRE) was calculated as the root-mean-square distances between point locations propagated using the CPD/LGCPD registration transformation versus using the gold standard displacement field (derived using a FE simulation or fluid registration for the cerebral and liver vessel images, respectively).

3 Results

An example of LGCPD versus CPD registration of the cerebral vessel is illustrated in Fig.1. The TRE results for all the cerebral vessel registrations are summarised in Table 1. The initial TREs for the vessel and the tumour before registration were 19.24 and 18.47 mm, respectively. Approximately 250 points were registered using the centreline representation, and 1000 using the surface representation. It can be seen from Table 1 that the LGCPD outperforms the CPD method in terms of registration accuracy. Furthermore, LGCPD registration using the centreline points was found to lead to a lower TRE for the tumour in all cases.

For the liver vessels, approximately 150 (250) and 300 (500) points were used for the source (target) centreline and surface representations, respectively. Fig.2 shows the result of registering these vessels. The TREs using the standard CPD method were 7.35 and 6.57mm using surface and centreline points, respectively, whereas these were improved to 5.51 and 2.60mm, respectively, by using LGCPD registrations. The initial TRE before registration was 14.12 mm. The TRE of a voxel as function of distance from the liver vessel surface is plotted in Fig.3. It was found that for any location with distance to the vessel surface smaller than 6.8mm, the TRE was smaller than 3mm. As the example shown in Fig.2, visual assessment of the registration results indicates incorrect correspondence found using the CPD method; in contrast,

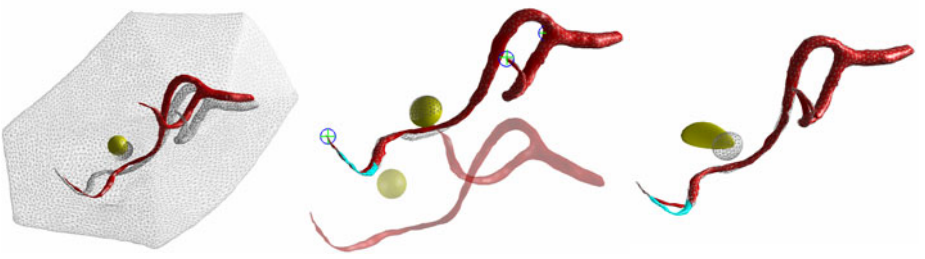


Fig. 1. Registration of a cerebral vessel (red), extracted from a patient MR image, to a deformed and corrupted version of itself (gray mesh). Left: The gray meshes represent the deformed vessel and tumour following a FE simulation. Middle: Example of the source vessel before and after registration using the LGCPD algorithm, shown in pink and red, respectively. Regions of the vessel removed to test the registration algorithm are shown in cyan, whilst the green crosses and blue circles indicate corresponding landmarks (bifurcations and ends of vessels). Right: Example of using the CPD registration algorithm without landmarks. It can be seen that, compared to the middle image, the alignment of the tumour and the ends of the longer branch are much less accurate.

Table 1. Summary of the TREs following registration of the cerebral vessels

TREs in mm (mean±SD)		**Missing Data**		**Noisy Data**		**Missing & Noisy Data**	
		Vessel	ROI	Vessel	ROI	Vessel	ROI
CPD	**Surface**	2.88±0.87	5.20±1.59	1.15±0.21	3.53±1.60	2.51±1.05	4.27±1.75
	Centreline	3.02±0.97	4.45±1.81	2.40±0.78	4.87±2.05	3.43±1.06	5.52±2.51
LGCPD	**Surface**	0.80±0.16	1.26±0.37	0.63±0.08	1.15±0.23	0.87±0.26	1.20±0.27
	centreline	0.58±0.18	0.54±0.12	1.02±0.18	1.05±0.33	1.12±0.22	1.10±0.33

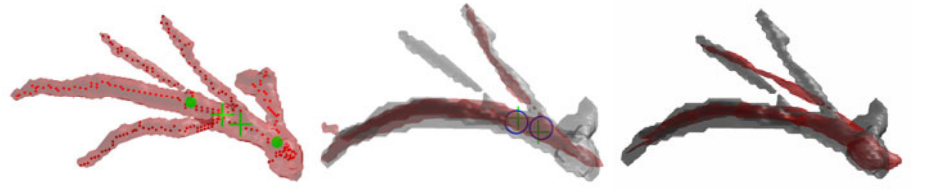


Fig. 2. Left: The right hepatic vein surface and centreline (shown in red) extracted from a deformed MR image of the liver. The bifurcations are denoted by green markers; the green crosses indicate the two landmarks used in the LGCPD registrations. Middle: example of a successful LGCPD registration showing the source vessel (red) aligned with the target vessel (gray). Corresponding landmarks (bifurcations) are shown as green crosses and blue circles. Right: Example of a CPD registration performed without using landmarks. It can be clearly seen that one source branch is aligned with the incorrect target branch.

the use of landmarks ensured that none of the LGCPD registrations produces such errors, both in the brain and liver cases.

Although the proposed LGCPD algorithm has the same computational complexity, and therefore speed, as the original CPD algorithm, it was found to execute faster in practice because fewer iterations were required to converge to a solution. In all the

experiments using our (unoptimised) implementation on a desktop PC with a 2.33GHz Intel® Core™ dual CPU processor and 3GB RAM, the centreline-based LGCPD registrations all converged within 1s. Registrations of the relatively larger number of vessel surface points were completed within 3s.

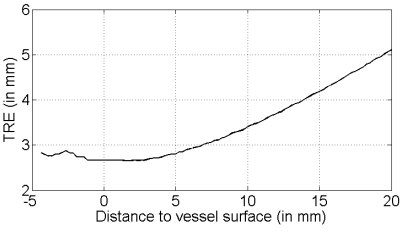


Fig. 3. Plot of the TREs of all voxels w.r.t. corresponding distances to the vessel surface (negative values indicate locations *inside* a vessel)

4 Discussion and Conclusions

In this paper, a new landmark-guided registration approach, based on the CPD algorithm, is proposed. The proposed method enables point landmarks to be incorporated in manner consistent with the original CPD formulation, resulting in an algorithm that is straightforward to implement, faster and more accurate than the original for vessel-based registration applications. Using centreline and surface point representation of vessels extracted from brain MRA and liver MR data, our results indicate that registrations can be performed within a few seconds, meaning that the method is highly appropriate for intraoperative applications, provided a method is employed that enables such representations to be generated in a similar timescale. A number of algorithms for automatic bifurcation identification and automatic vessel segmentation exist [16]. For the purpose of this study, all the landmarks were identified manually. One topic for further investigation is incorporation of a robust, automatic method to extract corresponding landmarks.

In this study, we used an unoptimised implementation of a well-known vesselness filter combined with simple thresholding and thinning to extract vessels and centreline points automatically. However, since all of these processes are parallelisable, significant speed-up in execution would be readily achievable using dedicated hardware. Furthermore, the algorithms could be optimised, for example, to limit the search space to regions near to vessels, or to utilise prior knowledge on likely vessel location and geometry form preoperative images or an initial slower registration, performed at the start of interventional procedure.

Acknowledgments. This work was supported by a UCL/UCLH Comprehensive Biomedical Research Centre New Initiative Project Grant (Ref. 96) and EPSRC Grant EP/F025750/1. Dean Barratt is funded by a Royal Academy of Engineering/EPSRC Research Fellowship. The Authors would like to thank Mr. Neil Kitchen and Dr. John Thornton from the National Hospital for Neurology & Neurosurgery, Dr. Freddy

Odille from UCL, and Prof. Tobias Schaefter from Imaging Sciences, King's College London for assistance in acquiring MR data for this study.

References

1. Aylward, S.R., Jomier, J., Weeks, S., Bullitt, E.: Registration and Analysis of Vascular Images. Int. J. Comput. Vision 55(2-3), 123–138 (2003)
2. Lange, T., Eulenstein, S., Hunerbein, M., Schlag, P.M.: Vessel-based non-rigid registration of MR/CT and 3D ultrasound for navigation in liver surgery. Computer Aided Surgery 8(5), 228–240 (2003)
3. Lange, T., Papenberg, N., Heldmann, S., Modersitzki, J., Fischer, B., Lamecker, H., Schlag, P.: 3D ultrasound-CT registration of the liver using combined landmark-intensity information. International Journal of Computer Assisted Radiology and Surgery 4(1), 79–88 (2009)
4. Reinertsen, I., Descoteaux, M., Siddiqi, K., Collins, D.L.: Validation of vessel-based registration for correction of brain-shift. Medical Image Analysis 11(4), 374–388 (2007)
5. Hipwell, J.H., Penney, G.P., McLaughlin, R.A., Rhode, K., Summers, P., Cox, T.C., Byrne, J.V., Noble, J.A., Hawkes, D.J.: Intensity-based 2-D-3-D registration of cerebral angiograms. IEEE Trans. Med. Imaging 22(11), 1417–1426 (2003)
6. Slomka, P.J., Mandel, J., Downey, D., Fenster, A.: Evaluation of voxel-based registration of 3-D power Doppler ultrasound and 3-D magnetic resonance angiographic images of carotid arteries. Ultrasound in Medicine & Biology 27(7), 945–955 (2001)
7. Penney, G.P., Blackall, J.M., Hamady, M.S., Sabharwal, T., Adam, A., Hawkes, D.J.: Registration of freehand 3D ultrasound and magnetic resonance liver images. Medical Image Analysis 8(1), 81–91 (2004)
8. Myronenko, A., Song, X.: Point Set Registration: Coherent Point Drift. IEEE Trans. on Pattern. Analysis. and Machine. Intelligence (2010) (in press) (pre prints)
9. Manniesing, R., Viergever, M., Niessen, W.: Vessel enhancing diffusion: A scale space representation of vessel structures. Medical Image Analysis 10(6), 815–825 (2006)
10. Frangi, A., Niessen, W., Vincken, K., Viergever, M.: Multiscale vessel enhancement filtering. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 130–137. Springer, Heidelberg (1998)
11. Homann, H.: Implementation of a 3D thinning algorithm. Insight Journal (2007)
12. Fang, Q., Boas, D.: Tetrahedral mesh generation from volumetric binary and gray-scale images. In: Proceedings of the Sixth IEEE International Conference on Symposium on Biomedical Imaging: From Nano to Macro, pp. 1142–1145. IEEE Press, Boston (2009)
13. Taylor, Z.A., Comas, O., Cheng, M., Passenger, J., Hawkes, D.J., Atkinson, D., Ourselin, S.: On modelling of anisotropic viscoelasticity for soft tissue simulation: Numerical solution and GPU execution. Medical Image Analysis 13(2), 234–244 (2009)
14. Crum, W.R., Tanner, C., Hawkes, D.J.: Anisotropic multi-scale fluid registration: evaluation in magnetic resonance breast imaging. Phys. Med. Biol. 50, 5153–5174 (2005)
15. Rijkhorst, E.J., Heanes, D., Odille, F., Hawkes, D., Barratt, D.: Simulating Dynamic Ultrasound using MR-derived Motion Models to assess Respiratory Synchronisation for Image-Guided Liver Interventions. In: Navab, N., Jannin, P. (eds.) IPCAI 2010. LNCS, vol. 6135, pp. 113–123. Springer, Heidelberg (2010)
16. Lesage, D., Angelini, E., Bloch, I., Funka-Lea, G.: A Review of 3D Vessel Lumen Segmentation Techniques: Models, Features and Extraction Schemes. Medical Image Analysis (2009)

Registration of CT Segmented Surfaces and 3-D Cardiac Electroanatomical Maps

Jingjing Gong[1], Yiyong Sun[2], and Zhaohua Chang[2]

[1] Biomedical Engineering, University of Shanghai for Science and Technology, 516 Jungong Road, Shanghai, 200093, China
[2] MicroPort Medical (Shanghai) Co. Ltd., Shanghai, 201203, China

Abstract. To improve the intra-operative image fusion performance in the ablation procedure of atrial fibrillation (AF) treatment, this paper presents a novel registration method for CT segmented surfaces and three-dimensional (3-D) cardiac electroanatomical maps. Random perturbation is introduced to deform the electroanatomical maps in the registration process. The magnitude of deformation automatically attenuates during iterations. Compared to the typical iterative closest point (ICP) algorithm that often converges to local minima, the proposed algorithm is much less sensitive to the initial transformation and able to move out of the local minima and converge to a solution with smaller registration error. Through experiments using both in vivo and simulation data, the results show significant improvements on the registration accuracy and success rate over the existing method being used in the clinical environment. The improved intra-operative registration results can help physicians easily navigate the catheter during the AF interventional procedures.

Keywords: intra-operative registration, electroanatomical maps, ICP, local minima.

1 Introduction

Atrial fibrillation (AF) is the most common cardiac arrhythmia [1]. During AF episodes, the normal electrical impulses generated from the sinoatrial node are overwhelmed by disorganized electrical impulses that originate in the atria and pulmonary veins, leading to conduction of irregular impulses to the ventricles. It results in irregular heartbeats which may occur in episodes lasting from minutes to weeks, or it could occur all the time over years. AF risk increases with age, with 8% of people over 80 having AF [2]. AF is often asymptomatic and is not in itself generally life-threatening, but may result in palpitations, fainting, chest pain or congestive heart failure. People with AF usually have a significantly increased risk of stroke [3,4].

AF may be treated with medications which either slow the heart rate or revert the heart rhythm back to normal. Synchronized electrical cardioversion may be used to convert AF to a normal heart rhythm. Surgical therapies may also be

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 70–77, 2010.

used to prevent recurrence of AF in certain individuals. In recent years, minimal invasive radiofrequency ablation has become a well-established therapy for the AF treatment. A catheter is inserted through a vein and radiofrequency energy is applied on the tip electrode to destroy abnormal electrical pathways in heart tissue or normal parts that are contributing to a cardiac arrhythmia. Besides the fluoroscopy, the physician also sees the catheter tip on the monitor, which is being tracked in three-dimensional (3-D) space. In a typical AF interventional procedure, the physician creates a 3-D electroanatomical map of the heart chamber by sampling a number of points on the endocardium to guide the subsequent ablations. An image integration module CARTOMERGE™ was introduced by Johnson & Johnson, which enables the fusion of the 3-D electroanatomical maps and the preoperative computed tomography (CT) or magnetic resonance images (MRI). Merging CT/MRI with the 3-D electroanatomical maps assists physicians more effectively to navigate catheter to the targeted points within the heart.

The registration accuracy of CT/MRI and the 3-D electroanatomical map has a direct impact on the AF interventional procedure results [5-10]. Registration of CARTOMERGE™ [11] is based on the iterative closest point (ICP) algorithm proposed by Besl and Mckay [12], which minimizes the distance between two point clouds by iteratively revising the rigid transformation. One of the problems of the ICP algorithm is that the solution is often trapped in local minima, especially when the source and target models have a large initial offset. User interaction is needed to initialize the transformation before applying the registration algorithm.

We propose a retrospective ICP (RICP) algorithm for registration of CT/MRI segmented surface and the 3-D electroanatomical map. The algorithm deforms the 3-D electroanatomical map by applying random displacement to the sampled electroanatomical points. During the iterations, the algorithm retrospects previous results and automatically decreases the deformation level. The method effectively overcomes the local minima problem and the final results give a smaller or at least equal registration error than the ICP algorithm. The improved intraoperative registration results can further help the physicians during the AF interventional procedures.

The rest of this paper is organized as follows. In Section 2, the ICP and stochastICP algorithm are briefly reviewed. Section 3 presents the RICP registration method of CT segmented surfaces and the 3-D electroanatomical maps. Section 4 shows the experimental results and the paper concludes in Section 5.

2 Review of the ICP and StochastICP Algorithms

The ICP algorithm is designed to rigidly register two point clouds, from a source model to a target model. It minimizes the sum of square distances with respect to the target points and their corresponding closest source points. First, an initial estimate of the transformation must be provided. Then for each point in the target model, the ICP algorithm finds the corresponding closest point on the

source model. These corresponding point pairs are used to calculate a translation vector and a rotation matrix that transform the source model towards the target model. This process is repeated until the sum of square distances between two point clouds is below a predetermined value.

Since the ICP algorithm is prone to be trapped in local minima, the stochastICP algorithm was proposed by Penney in 2001 [13], which applies random Gaussian noise to perturb the coordinates of one of the two models. The transformation is expressed by 6-tuple parameters of three angles and three offsets. If the maximal change of all parameters between the current iteration and one of the previous sets of iterations is within a threshold, the algorithm reduces the perturbation magnitude. The threshold, which is dependent on dataset, is empirically chosen as a fraction of the perturbation magnitude. Although the proposed RICP algorithm is also based on random perturbation concept, we retrospect to a previous iteration with the least registration error to guide the perturbation magnitude reduction.

3 Registration Algorithm of CT Segmented Surfaces and the 3-D Cardiac Electroanatomical Maps

The proposed RICP registration algorithm in this paper can make that the registration error is smaller than or at least equal to the ICP algorithm. We apply this algorithm to the rigid registration of CT segmented surfaces and the 3-D cardiac electroanatomical maps.

We define CT segmented surface as the source model and the 3-D cardiac electroanatomical map as the target model. Assume the target model consists of m points and the source model consists of n points, which are denoted respectively by $p = \{p_i | i = 1, \cdots, m\}$ and $q = \{q_j | j = 1, \cdots, n\}$ (p_i and q_j are considered as 3×1 column vectors). Generally, n is much larger than m since the CT segmented surface has tens of thousands of vertices and the 3-D cardiac electroanatomical map usually has less than a hundred points. Two models are in different coordinate systems with a transformation

$$p = Rq + T, \tag{1}$$

where T is the translational offset and R is the rotation matrix.

The registration error for each point p_i in the target model is

$$\|e_i\| = \|p_i - (Rq_i + T)\|, \tag{2}$$

where q_i is the corresponding closest point of p_i searched within q using a KD-tree algorithm. Thus, we obtain m point pairs and minimize the average of sum of the square errors

$$E = \frac{1}{m} \sum_{i=1}^{m} ||e_i||^2. \tag{3}$$

Before corresponding point pairs are searched, we impose a random perturbation n_i to each point p_i,

$$p_i' = p_i + n_i, \tag{4}$$

so that the target model is deformed, where n_i is a 3×1 column vector whose components have a normal distribution $N(0, \sigma^2)$.

For points $p' = \{p'_i | i = 1, \cdots, m\}$ on the deformed target model, we search the corresponding nearest points $q' = \{q'_i | i = 1, \cdots, m\}$ on the source model. We define centroids of p' and q' by

$$\bar{p} = \frac{1}{m} \sum_{i=1}^{m} p'_i \text{ and } \bar{q} = \frac{1}{m} \sum_{i=1}^{m} q'_i. \tag{5}$$

Let us denote $p''_i = p'_i - \bar{p}$ and $q''_i = q'_i - \bar{q}$. We calculate the 3×3 matrix

$$M = \sum_{i=1}^{m} q''_i (p''_i)^t, \tag{6}$$

and find the singular value decomposition (SVD) of M

$$M = UDV^t. \tag{7}$$

Here, the superscript t denotes matrix transposition, U and V are 3×3 orthonormal matrices, and D is a diagonal matrix. Then we calculate R and T by

$$R = VU^t \text{ and } T = \bar{p} - R\bar{q}. \tag{8}$$

Finally, the transformation between the two models is

$$Z = \begin{pmatrix} R & 0 \\ T & 1 \end{pmatrix}. \tag{9}$$

We define the initial registration error as E_0 and the initial transformation as Z_0. We apply the transformation matrix Z to the source model in order to transform it towards the target model. Due to the random perturbation, the registration error could either increase or decrease. We repeat above procedure L times, calculate and record all the transformations $\{Z_i | i = 1, \cdots, L\}$ and the corresponding registration errors $\{E_i | i = 1, \cdots, L\}$. Then we get $E_{\min} = \min\{E_1, E_2, \cdots E_L\}$ and the corresponding transformation $Z_{\min}$. If $E_{\min} \leqslant E_0$, we replace Z_0 with $Z_{\min}$ as the initial transformation for the next L times iterations, otherwise we keep Z_0 as the initial transformation. At the same time, the control parameter of random perturbation is automatically attenuated as $\sigma = \sigma/\sqrt{2}$.

The iteration stops when the registration error between the two models is below a predetermined threshold. The algorithm also stops if it reaches a predetermined maximum number of iterations. For different models, the algorithm starts with an arbitrarily large initial value of σ. If the initial transformation is already the best, the registration result would remain the same. Typically, the RICP algorithm gives a better or at least the same solution as the ICP algorithm.

The proposed rigid registration method has a much larger domain of convergence and can effectively move out of the local minima. The registration accuracy and success rate are higher than the ICP algorithm.

4 Experimental Results

All experiments in the paper are performed on a 2.0 GHz Pentium 4 computer with 1 GB of RAM. The criterion for stopping the iteration is that the registration error is smaller than 0.1 mm. The initial perturbation parameter σ is defined to be 50 and L is defined as 100.

4.1 The in Vivo Models Study

The left atrium segmented surfaces in the experiments are generated from the CT datasets of AF patients, using a threshold-based segmentation algorithm and 3-D punching tool of a 3-D image analysis software. The 3-D cardiac electroanatomical maps are obtained using the ablation catheter tip to pick a number of points on the endocardium of patient's left atrium. We use CT segmented left atrium surfaces and the electroanatomical maps of four patients. For each patient data, we randomly generate 800 transformations, each of which consists of a translational offset within the scope of 100 mm in x, y, and z axes directions and a rotation deviation within the scope of 20 ° around x, y, and z axes. These transformations are applied to each pair of models as the initial registration.

Table 1. The registration results of different initial translations and rotations

Patient	Ave. Error (mm)		Std. Dev.		Success Rate		Computation Time (s)	
Data	ICP	RICP	ICP	RICP	ICP	RICP	ICP	RICP
1	4.30	1.65	3.14	0.04	53.6%	99.6%	0.16	6.82
2	2.37	1.91	1.06	0.12	79.3%	99.1%	0.14	5.51
3	5.04	1.79	3.59	0.06	42.6%	99.8%	0.07	3.15
4	1.92	1.66	0.58	0.05	83.6%	99.8%	0.15	6.60

Table 2. The registration results of different initial rotations

Initial	Ave. Error (mm)		Std. Dev.		Success Rate	
Rotation	ICP	RICP	ICP	RICP	ICP	RICP
30 °	4.27	2.18	2.75	0.45	32%	96%
45 °	5.92	1.75	3.72	0.28	26%	88%
60 °	5.89	1.68	3.52	0.16	14%	37%

Table 3. The registration results of the simulation data

Simulation	Ave. Error (mm)		Std. Dev.		Max Error (mm)		Min Error (mm)	
Data	ICP	RICP	ICP	RICP	ICP	RICP	ICP	RICP
1	4.04	2.49	2.12	0.76	13.86	11.92	0.14	0.05
2	3.38	2.47	1.92	0.94	11.20	8.86	0.08	0.06
3	3.98	2.35	1.88	0.82	12.66	7.73	0.14	0.07
4	3.90	2.74	1.68	0.69	13.81	9.58	0.13	0.21
5	3.74	1.91	2.68	1.02	11.14	6.52	0.17	0.06

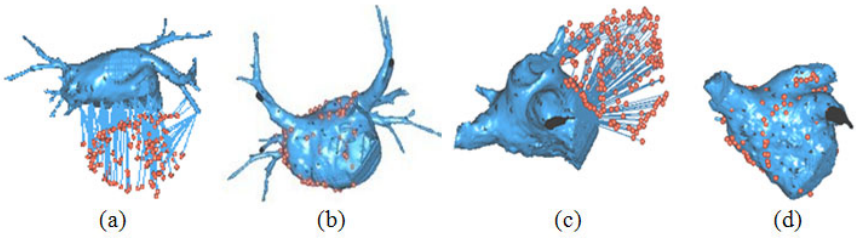


Fig. 1. Image fusion results. (a) before registration of patient data 1, (b) after registration of patient data 1, (c) before registration of patient data 2, (d) after registration of patient data 2

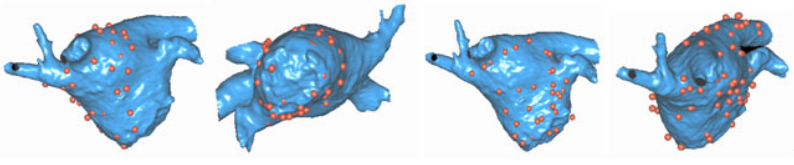


Fig. 2. The registration and image fusion results of non-uniform distribution models

Table 1 shows the average registration error of 800 registration attempts and success rate for the ICP and RICP algorithms. Success rate is defined as the percentage of registrations from 800 initial positions in which the registration error is smaller than 2 mm.

The registration errors of the RICP algorithm are significantly smaller than the ICP algorithm. The success rates are also much higher than the ICP algorithm. It is not sensitive to the initial positions compared to the ICP algorithm. On the other hand, the RICP is more time consuming since it retrospectively seeks the optimal position in every L iterations. In Table 1, we list the computation time for all datasets. Since the user does not need to perform manual alignment, the increased computation time is acceptable to automatically get a more accurate registration.

Since the registration result is more sensitive to the initial rotation compared to the initial translation, we perform the experiments by applying larger initial rotations. For 100 times, the source model is given initial positions with a fixed 100 mm translational offset and different rotations of 30 °, 45 °, and 60 °. Table 2 shows the average registration error and success rate of 100 registration attempts for one pair of models. Although the success rate drops when the initial angle deviation increases, the RICP still performs better than the ICP algorithm. Figure 1 shows the image fusion results of CT segmented surface of left atrium and the electroanatomical map represented by a point cloud. The lines show the connection of points on the electroanatomical map and the corresponding nearest points on the CT segmented surface.

4.2 The Simulation Study

For the simulation study, the left atrium segmented surface is generated from a CT dataset of a patient. The left atrium anatomical map is obtained using the catheter to pick a number of points on a transparent plastic left atrium model which is made by rapid prototyping of the segmented surface. We sample the model for five times and create five simulation data of the left atrium. Table 3 shows the registration results, including the average registration error, standard deviation, maximum error, and minimum error.

Figure 2 shows image fusion results of four pairs of models, in which the 3-D mapping points distribute only on one side of the simulation data. The result shows that the RICP algorithm performs well for non-uniform distribution models.

5 Conclusions

We have presented a novel rigid registration method for CT segmented surfaces and the 3-D cardiac electroanatomical maps. From the results using in vivo and simulation data, the proposed RICP registration algorithm can effectively overcome the local minima problem. It is more robust and accurate than the ICP algorithm. In addition, it performs well on the non-uniform distribution models. The improved intra-operative registration results can greatly help the physicians during the AF interventional procedures. Future work will focus on thorough validation using more datasets from the clinical environment of the atrial fibrillation interventional procedures.

References

1. Wyndham, C.R.C.: Atrial Fibrillation: the Most Common Arrhythmia. Texas Heart Institute Journal 27(3), 257–267 (2000)
2. Kannel, W.B., Benjamin, E.J.: Status of the Epidemiology of Atrial Fibrillation. Medical Clinics of North America 1(92), 17–40 (2008)
3. Benjamin, E.J., Wolf, P.A., D'Agostino, R.B., Silbershatz, H., Kannel, W.B., Levy, D.: Impact of Atrial Fibrillation on the Risk of Death: the Framing Ham Heart Study. Circulation 10(98), 946–952 (1998)
4. Wattigney, W.A., Mensah, G.A., Croft, J.B.: Increased Atrial Fibrillation Mortality: United States, 1980-1998. American Journal of Epidemiology 9(155), 819–826 (2002)
5. Mikaelian, B.J., Malchano, Z.J., Neuzil, P., Weichet, J., Doshi, S.K., Ruskin, J.N., Reddy, V.Y.: Integration of 3-Dimensional Cardiac Computed Tomography Images with Real-Time Electroanatomic Mapping to Guide Catheter Ablation of Atrial Fibrillation. Circulation 112, e35–e36 (2005)
6. Kistler, P.M., Rajappan, K., Jahngir, M., Earley, M.J., Harris, S., Abrams, D., Gupta, D., Liew, R.: The Impact of CT Image Integration into an Electroanatomic Mapping System on Clinical Outcomes of Catheter Ablation of Atrial Fibrillation. Journal of Cardiovascular Electrophysiology 10(17), 1093–1101 (2006)

7. Bella, P.D., Fassini, G., Cireddu, M., Riva, S., Carbucicchio, C., Giraldi, F., Maccabelli, G., Trevisi, N., Moltrasio, M., Pepi, M., Galli, C.A., Andreini, D., Ballerini, G., Pontone, G.: Image Integration Guided Catheter Ablation of Atrial Fibrillation: a Prospective Randomized Study. Journal of Clinical Cardiology 20, 258–265 (2009)
8. Xu, J., Chen, Y., Li, H., He, B.: Successful Ablation of Complicated Supraventricular Tachycardia under the Carto System Guidance in One Patient. Chinese Journal of Cardiology 37(11), 1044–1046 (2009)
9. Bian, C., Ma, J., Jiang, J., Yao, S., Lv, X., Wang, J.: Typical Atrial Flutter Radiofrequency Catheter Ablation under the Carto System Guidance in Nine Patients. Journal of Electrocardiology (China) 28(5), 346–347 (2009)
10. Bertaglia, E., Bella, P.D., Tondo, C., Proclemer, A., Bottoni, N., Ponti, R.D., Landolina, M., Bongiorni, M.G., Cor, L., Stabile, G., Russo, A.D., Verlato, R., Mantica, M., Zoppo, F.: Image Integration Increases Efficacy of Paroxysmal Atrial Fibrillation Catheter Ablation: Results from the CartoMergeTM Italian Registry. EP-Europace 8(11), 1004–1011 (2009)
11. Sun, Y., Azar, F.S., Xu, C., Hayam, G., Preiss, A., Rahn, N., Sauer, F.: Registration of High-Resolution 3D Atrial Images with Electroanatomical Cardiac Mapping: Evaluation of Registration Methodology. In: Proc. SPIE, vol. 5744, pp. 299–307 (2005)
12. Besl, P.J., McKay, N.D.: A Method for Registration of 3-D Shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence 14(2), 239–256 (1992)
13. Penney, G.P., Edwards, P.J., King, A.P., Blackall, J.M., Batchelor, P.G., Hawkes, D.J.: A Stochastic Iterative Closest Point Algorithm (stochastICP). In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 762–769. Springer, Heidelberg (2001)

Coronary Motion Estimation from CTA Using Probability Atlas and Diffeomorphic Registration

Dong Ping Zhang[1], Laurent Risser[1,2], François-Xavier Vialard[2]
Philip Edwards[1], Coert Metz[3], Lisan Neefjes[4]
Nico Mollet[4], Wiro Niessen[3], and Daniel Rueckert[1]

[1] Department of Computing, Imperial College London, London, UK
[2] Institute for Mathematical Science, Imperial College London, London, UK
[3] Dept. of Medical Informatics and Radiology, Erasmus MC, Rotterdam, NL
[4] Dept. of Radiology and Cardiology, Erasmus MC, Rotterdam, NL

Abstract. In this paper, we present a method for coronary artery motion estimation from 4D cardiac CT angiogram (CTA) data sets. The proposed method potentially allows the construction of patient-specific 4D coronary motion model from pre-operative CTA which can be used for guiding totally endoscopic coronary artery bypass surgery (TECAB). The proposed approach consists of three steps: Firstly, prior to motion tracking, we form a coronary probability atlas from manual segmentations of the CTA scans of a number of subjects. Secondly, the vesselness response image is calculated and enhanced for end-diastolic and end-systolic CTA images in each 4D sequence. Thirdly, we design a special purpose registration framework for tracking the highly localized coronary motion. It combines the coronary probability atlas, the intensity information from the CTA image and the corresponding vesselness response image to fully automate the coronary motion tracking procedure and improve its accuracy. We performed pairwise 3D registration of cardiac time frames by using a multi-channel implementation of the Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework, where each channel contains a given level of description of the registered shapes. For validation, we compare the automatically tracked coronaries with those segmented manually at end-diastolic phase for each subject.

1 Introduction

As one of the leading causes of death worldwide, coronary artery disease occurs due to the failure of the blood circulation to supply adequate oxygen and nutrition to cardiac tissues. It is typically caused by the excessive accumulation of atheromatous plaques and fatty deposits within certain regions of the arteries which restrict the blood flow. To treat this disease, arteries or veins grafted from the patient's body are used to bypass the blockages and restore the supply to the heart muscle. Conventional bypass surgery requires invasive sternotomy and the use of a cardiopulmonary bypass machine, which leads to

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 78–87, 2010.

long recovery period for the patient and has high infectious potential. Using image-guided robotic surgical system, totally endoscopic coronary artery bypass (TECAB) surgery techniques have been developed to allow clinicians to perform bypass surgery off-pump with three pin-hole incisions in the chest cavity, through which two robotic arms and one stereo endoscopic camera are inserted. However, 20-30% conversion rates from TECAB surgery to conventional invasive surgical approach [1,2] have been reported due to the vessel misidentification and mislocalization caused by the restricted field of view of the stereo endoscopic images. To reduce this conversion rate and facilitate the TECAB procedure, we aim to construct a patient-specific 4D coronary artery motion model from preoperative cardiac CTA sequence. The main challenge of constructing this motion model is to follow the motion and deformation of the coronaries from end-systolic to end-diastolic phase accurately. In this paper, we propose a pyramid registration framework to achieve this. Finally, through temporally and spatially aligning the coronary motion model with the intraoperative endoscopic views of the patient's beating heart, this work has the potential to assist the surgeon to identify and locate the correct coronaries during the robotically-controlled TECAB procedures.

1.1 Related Work

Recent advances in computed tomography of coronary arteries [3] have attracted several studies on using CTA for coronary artery disease diagnosis and surgical planning. Extensive reviews on coronary artery segmentation are given in Schaap *et al.* [4] and Lesage *et al.* [5]. Although coronary artery segmentation has been well studied, particularly in the Grand Challenge of Coronary Artery Centerline Extraction [6], constructing motion models of coronaries from pre-operative CTA sequences to assist the diagnosis and surgery is a topic which has received less attention. The work in this paper differs significantly from 3D vessel segmentation approaches: our aim is not the extraction of the coronaries in single-phase high-quality CTA datasets as in [6], but instead estimating the coronary motion from end-systolic to end-diastolic phase in multi-phase cardiac CTA sequence.

Previously, Shechter *et al.* [7,8] tracked coronary artery motion in a temporal sequence of biplane X-ray angiography images. In their approach, a 3D coronary model is reconstructed from extracted 2D centrelines in end-diastolic angiography images. The deformation throughout the cardiac cycle is then recovered by a registration-based motion tracking algorithm. The limitation of this approach is that 3D reconstruction of the coronary from 2D X-ray images is required. An alternative approach for the extraction of the coronaries from cardiac CTA has been proposed by Metz *et al.* [9]: Here the coronaries are manually or semi-automatically identified at one time frame and then tracked throughout the cardiac cycle using non-rigid registration of the multi-phase cardiac CTA images. The restriction of this approach is that highly localized motion of the coronaries cannot be fully recovered by the motion tracking of the entire heart, particularly for right coronary artery as shown in Fig. 2 and Fig. 3 in [9].

In this paper, we present a novel approach for automatic segmentation of the coronary lumen and motion tracking of the coronaries from end-systolic to end-

diastolic phase in 4D cardiac CTA images. This is achieved by using tensor voting to improve the connectivity of the coronary vessels in vesselness response image and a special-purpose registration framework for coronary motion tracking. The advantages of the proposed approach here are two folds. Firstly, it contains less manual pre-processing procedures, since no 3D reconstruction from X-ray images is required in order to perform 3D coronary motion tracking as in [7,8] and no manual segmentation or user-iteraction is required for patient-specific coronary motion modeling. Secondly, by combining greylevel CTA image, its vesselness response image and coronary probability atlas in the proposed registration framework, this approach provides a robust estimation of the coronary motion in 4D CTA.

2 Method

In order to augment the intraoperative images acquired with stereo-endoscope during TECAB procedure, a patient-specific 4D motion model of coronaries is constructed from the pre-operative dynamic CTA sequence. This is achieved by forming coronary probability atlas and using multi-channel LDDMM registration [10,11]. The formalism of LDDMM makes large diffeomorphic (smooth and invertible) transformations possible when registering two shapes. Contrary to alternative methods [12,13], the LDDMM formalism is designed to compute shape deformations that are geodesics. An optimal flow of deformation is then estimated between the source and the target images according to a regularization metric and a similarity measure. Being able to follow the coronary motion using this formalism then offers new possibilities for the statistical characterization of this motion. However, in this context and for coronary motion tracking from CTA sequences, mono-channel registration on the grey level images does not capture accurately the highly localized coronary motion which is surrounded by large anatomical structures . The method in [10] has previously been extended to treat multi-channel images [11]. Our methodological contribution is then to use multi-channel registration, where each channel contains a level of description of the registered shapes: (1) original images, (2) coarse but robust extractions of the coronary vessels and (3) highly smoothed original images. We then denote our approach as Multi Level- or ML-LDDMM.

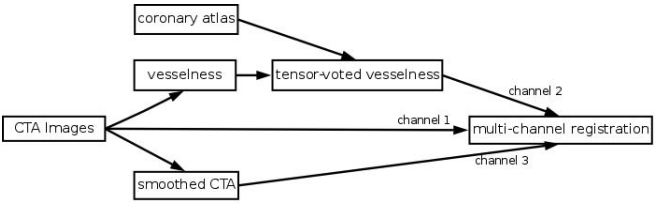


Fig. 1. Overview of the proposed coronary motion estimation method

By using tensor voting and extracting the largest connected components, the coronary lumen is segmented from the post-processed vesselness response image at end-systolic phase for each 4D CTA sequence. The segmented coronary lumen is then transformed to the end-diastolic time frame according to the deformation from ML-LDDMM registration. The estimated coronary lumen at end-diastolic is then compared to manual tracking of the coronaries at end-diastolic phase. Fig. 1 shows an overview and illustrates the connections between the components in our proposed algorithm.

2.1 Coronary Atlas Construction Using Affine Registration

Chillet *et al* [14] presented a method for forming brain and liver vascular atlases using a vessel-to-image affine registration method. Here a 3D probability atlas $\mathcal{A}$ containing left anterior descending (LAD), left circumflex artery (LCX) and right coronary artery (RCA) is constructed from the manually marked centerlines in a group of single-phased 3D CTA scans acquired from 26 patients. One subject's CTA scan is chosen as reference $\mathcal{R}$ and the rest of the CTA scans are affine aligned with $\mathcal{R}$. The manually segmented centerlines are transformed according to the resulting affine transformation in order to match with the centerline from the reference subject. All the transformed centerlines and the corresponding reference centerline are then wrapped as a tubular structure with a pre-defined radius and blurred with Gaussian kernel. For each coronary artery, the blurred tubular structures are averaged together using a Gaussian kernel to create the atlas for that branch. The process is repeated for all three coronary arteries, LAD, LCX and RCA in order to form a probability atlas for coronary tree.

By affine alignment of the reference image $\mathcal{R}$ used for creating the probability atlas and the end-diastolic phase of a new patient CTA sequence, this atlas $\mathcal{A}$ is affinely registered to the patient dataset to create patient-specific coronary mask $\mathcal{M}_{ED}$. For each patient CTA sequence, the 3D mask $\mathcal{M}_{ED}$ is then warped to create a mask $\mathcal{M}_{SD}$ corresponding to end-systolic phase by non-rigid registration [15] of the CTA images. Each patient-specific coronary mask is then used to select the coronary artery region from its corresponding vessel response image and incorporated in the ML-LDDMM registration.

2.2 Multi-scale Vessel Enhancement Diffusion and Filtering

Prior to the coronary artery segmentation, vessel enhancing diffusion [16] is used to improve the visibility of the coronaries in CTA images. To reduce the effect of the presence of inhomogeneous background (e.g. air and tissue mixed region) or irrelevant neighboring structures (e.g. bone or metal implant), multiple thresholds are selected automatically by 4D multi-level thresholding extended from Otsu's method [17] for each 4D data set. The intensities of the background voxels are increased so that they match the average myocardial intensity level. Voxels with intensities above the upper threshold level that represents bone structure are assigned average myocardium intensity value.

We perform a coarse segmentation of the coronary arteries in the CTA image using a multiscale Hessian-based vessel enhancement filter [18]. The filter

utilizes the 2nd-order derivatives of the image intensity after smoothing (using a Gaussian kernel) at multiple scales to identify bright tubular-like structures with various diameters. The six second-order derivatives of the Hessian matrix at each voxel are computed by convolving the image with second-order Gaussian derivatives at pre-selected scales. Assuming a 3D image function $I(x)$, the Hessian matrix at a given voxel x at scale σ is denoted as $H_\sigma(x)$. A vesselness term $V(x)$ is defined as in Frangi *et al.* [18] and is based on the eigenvalues and eigenvectors of $H_\sigma(x)$. The vesselness response is computed at a range of scales, exponentially distributed between σ_{min} and σ_{max} . The maximum response with the corresponding optimal scale is obtained for each voxel of the image. The vesselness image is then constructed using the maximum response of each voxel as the intensity value.

2.3 Tensor Voting

Due to artifacts introduced in CT reconstruction and the low signal-to-noise ratio in certain phases caused by the application of ECG pulsing windows to reduce the radiation dose for the patient [19], segmenting the vessels directly from the vesselness images does not provide satisfactory results. To alleviate those effects, we propose to use a post-processing step of tensor voting [20,21] to enhance the extraction of the coronary vessels. Tensor voting was initially developed to reconstruct shapes from point clouds but was also shown efficient to recover volumes, surfaces and curves from noisy images. Here it is adapted to fill discontinuities in vesselness response image.

Consider a vesselness image $I_v \in \Omega$. We associate a tensor-valued image $\boldsymbol{TV} \in \Omega$ to I_v. Each voxel of $\boldsymbol{TV}$ is a 3×3 matrix that allows the local communication between the noisy data of I_v. Using the framework of [21], each non-null voxel x_i of I_v is considered as an island token i, where $i \in \{1, ..., N\}$. A token is then here a point being potentially in a curve according to the vesselness image. This token generates a tensor field $\boldsymbol{TV}_i$ around x_i. For $x \in \Omega$, $\boldsymbol{TV}_i(x)$ is computed as the tensor product of a vector $\boldsymbol{W}_i(x)$ with itself. This vector is defined as:

$$\boldsymbol{W}_i(x) = e^{-\frac{d(x_i,x)^2}{\delta^2}} \frac{x_i x}{d(x_i, x)}, \tag{1}$$

where $d(x_i, x)$ is the Euclidian distance between x_i and x and δ is a scale at which the structures are recovered. In our computations, we set δ to values slighly higher than the typical radius of the vessels. Note that in practice, we only compute $\boldsymbol{TV}_i(x)$ in a bounding box in which its values are not negligible. The communication between the island tokens is then performed in $\boldsymbol{TV}$ by using:

$$\boldsymbol{TV}(x) = \sum_{i=1}^{N} \boldsymbol{TV}_i(x), \quad \forall x \in \Omega. \tag{2}$$

By using the saliency map to a curve: $L(x) = \lambda_1(x) - \lambda_2(x)$, where $\lambda_1(x)$ and $\lambda_2(x)$ are the two highest eigenvalues of $\boldsymbol{TV}(x)$, we finally measure how the point x fits into a curve according to its neighborhood. This map enhances the

coronary vessels even where the contrast between the vessels and the surrounding tissues is low. Note that this simple interpretation of the tensor voting generates a small ghosting effect around the vessel centerlines. So $L(x)$ is smoothed with a Gaussian filter of standard deviation $\delta/2$.

2.4 Multi level LDDMM Registration

Image registration is performed using three channels ($C_{1,S}$, $C_{2,S}$, $C_{3,S}$) and ($C_{1,T}$, $C_{2,T}$, $C_{3,T}$) computed from the source image I_S and the target image I_T using the pipeline summarized in Fig. 1 and explained in the previous subsections. The first channel C_1 contains the original images I_S and I_T. The second one contains the enhanced vesselness images L_S and L_T, calculated in subsection 2.3. Finally, the third channel represents the original images smoothed by a large gaussian filter (5mm in our calculations). These channels are a multi-level representation of the registered shapes. The channel C_3 pushes the source image to the target image without any consideration of the details, so the global motion of the large components of the CTA images is tracked. The channel C_2 has a similar role, but only focuses on the coronary region. Finally, the channel C_1 is complementary to C_2 and C_3 since it takes image details into account.

A multi-level strategy, as in Fig. 2, is adopted to register the images efficiently. In the first level, an initial estimation of optimal deformation is obtained using the smoothed images C_3, then this estimation is refined using registration based on the second channel C_2. Finally by introducing channel C_1, all the information is taken into account for the finest estimation of deformation. This strategy is particularly robust. For the first two levels, it can be performed on down-sampled images speedily without losing much accuracy.

The registrations are performed using the LDDMM framework [10,11]. In this framework, the images are deformed through time dependent diffeomorphic transformations ϕ_t, $t \in [0, 1]$, which are defined by a time dependent velocity field

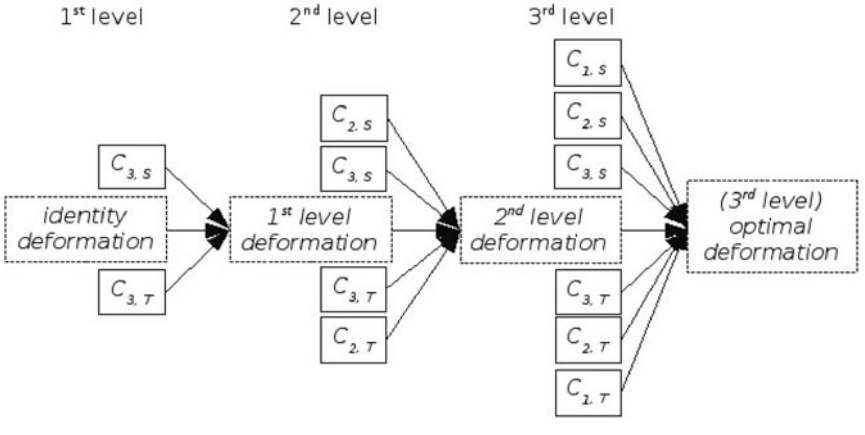


Fig. 2. Multi-level strategy of the registration

v as follows: $\partial_t \phi_t = v_t(\phi_t)$, where ϕ_0 is the identity deformation and $t \in [0,1]$. The velocity field v_t deforms the image coordinates at time t and ϕ_t is the induced deformation. For notational convenience, we introduce $\phi_{t,s} \doteq \phi_s \circ \phi_t^{-1}$. Contrary to [10], where the similarity measure is computed directly from I_S and I_T, the images are indirectly compared here using the information contained in the channels. The energy we minimize, as a function of v, is then:

$$E(v) = \int_0^1 \frac{1}{2}||v_t||_V^2 \mathrm{d}t + \sum_{i=1}^{3} \left(\alpha_i ||C_{i,S} \circ \phi_1^{-1} - C_{i,T}||_{L^2}^2\right), \tag{3}$$

where $\alpha_i \in [0,1]$, controls the weight of the channel i. These weights are tuned to have a similar influence of each channel. In the energy of the deformations, the time dependent velocity field v is assumed to lie in $L^2([0,1],V)$, where V is a Hilbert space of vector fields. Underlying this space, there exists a smooth matrix valued kernel K that controls the spatial correlation of the deformations. The minimization of the energy is described hereafter.

We denote $D_t^{i,S} = C_{i,S} \circ \phi_{t,0}$ and $D_t^{i,T} = C_{i,T} \circ \phi_{1,t}$. The Jacobian of $\phi_{t,1}$ at time t is also noted $|D\phi_{t,1}|$. The minimization of the variational problem of Eq. 3 is performed by using a steepest gradient descent approach. Practical resolution then involves the iterative use of the gradient of the functional E at time t:

$$\nabla_v E_t = v_t - K \star \left(|D\phi_{t,1}| \sum_{i=1}^{3} (\alpha_i \nabla D_t^{i,S} (D_t^{i,S} - D_t^{i,T})) \right), \tag{4}$$

where $\star$ is the convolution operator. In our computations, we used an isotropic Gaussian kernel K of standard deviation 10mm. Such kernel offers a good balance between enough spatial correlation to ensure vessel radius preservation and enough spatial flexibility to properly match curved vessels.

3 Results and Evaluation

The coronary atlas is constructed from the CTA scans of 26 patients. The proposed motion estimation algorithm is tested in 4D CTA sequences from eight subjects. For each subject, we segmented the coronary lumen automatically and performed a thinning algorithm to get the coronary tree centerlines from both

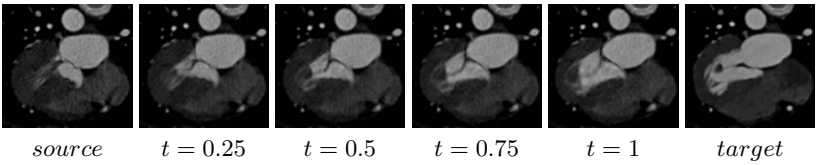


Fig. 3. ML-LDDMM registration results. From left to right, source image I_S, intermediate transformed results ($t = 0.25, 0.5, 0.75$), final result ($t = 1$) and target image I_T

end-systolic (ES) and end-diastolic (ED) phases. The centerlines in ES phase are then deformed according to the optimal deformation from ML-LDDMM registration to estimate the coronary positions in ED phase.

In order to quantitatively measure the accuracy of the motion tracking algorithm, we manually marked the centerlines for the ED and ES phases in these eight CTA sequences. We compare the transformed centerlines with the manually marked ones in ED phase as in the third row of Table 1. The distance between the manual segmentation U and the automatic tracked coronaries W in each time frame is measured as:

$$D(W, U) = \frac{1}{N_W} \sum_{i=1}^{N_W} \|w_i - l(w_i, U)\|_2 + \frac{1}{N_U} \sum_{j=1}^{N_U} \|u_j - l(u_j, W))\|_2 \quad (5)$$

where N_W and N_U are the number of points representing vessel W and vessel U correspondingly. For each point $m_i \in W$, $l(m_i, U)$ calculates the closest point to w_i on the automatically extracted vessel U. Similarly, for each point $u_j \in U$, $l(u_j, W)$ defines the closest point to u_j on the vessel W. As comparison, we

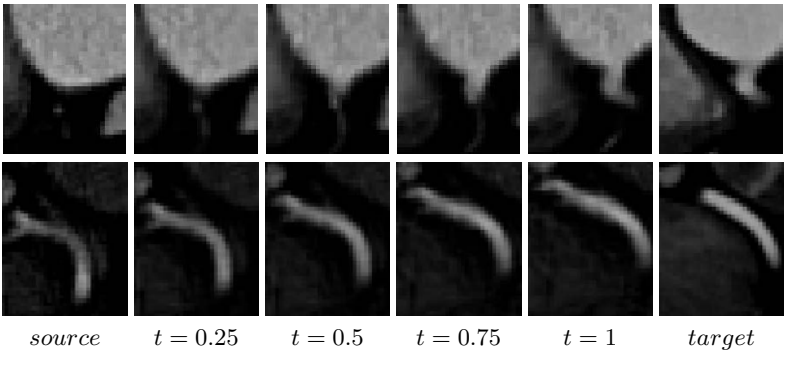


Fig. 4. Registration results. The top row shows starting position of RCA. From left to right, the source image, the intermediate transformed source images ($t = 0.25, 0.5, 0.75$), the final result ($t = 1$) and the target image. Similarly, the bottom row shows the mid-section region of RCA deforming from source to match with target. Note that in each row all images were taken in the same region of interest.

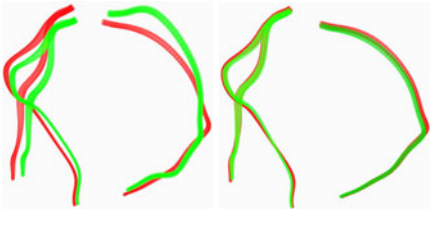


Fig. 5. Coronary artery lumen (P6). Left: segmented coronary artery lumen at ES (green) and ED (red) phases. Right: estimated coronary lumen at ED phase (green) compared with segmented lumen at ED (red).

Table 1. Coronary displacement and error of motion estimation

	P1	P2	P3	P4	P5	P6	P7	P8
Displacement (mm)	8.18	7.63	8.36	7.79	8.62	10.05	8.09	7.88
Estimation error (mm)	1.79	1.26	1.43	1.25	0.69	1.19	1.35	1.04
Compensation (%)	78.12	83.49	82.89	83.95	92.00	88.16	83.31	86.80

also present the natural displacement of coronary tree from ES to ED as in the second row in Table 1. The fourth row shows the percentage of coronary displacement that has been compensated by our motion estimation method. The results show that ML-LDDMM registration based motion estimation has performed robustly and accurately. By automatically segmenting the coronary artery and tracking the coronaries from ES to ED in CTA sequences, the patient-specific coronary model and motion estimation are performed robustly in all eight testing subjects.

4 Discussion

We have presented a novel approach for patient-specific coronary artery segmentation and motion estimation from dynamic cardiac CTA sequences that significantly improves the robustness of motion tracking and eliminates the manual interaction. The proposed method has been tested on the clinical CTA datasets acquired from eight subjects. By segmenting coronaries and tracking their motion from pre-operative cardiac images and aligning this motion with the series of 2D endoscopic images capture during the operation, we aim to assist the surgical planning and provide image guidance in robotic-assisted totally endoscopic coronary artery bypass (TECAB) surgery. Through this work, we expect to reduce the conversion rate from TECAB to conventional invasive procedures.

References

1. Mohr, F.W., Falk, V., Diegeler, A., Walther, T., Gummert, J.F., Bucerius, J., Jacobs, S., Autschbach, R.: Computer-enhanced "robotic" cardiac surgery: Experience in 148 patients. J. Thorac. Cardiovasc. Surg. 121(5), 842–853 (2001)
2. Dogan, S., Aybek, T., Andressen, E., Byhahn, C., Mierdl, S., Westphal, K., Matheis, G., Moritz, A., Wimmer-Greinecker, G.: Totally endoscopic coronary artery bypass grafting on cardiopulmonary bypass with robotically enhanced telemanipulation: Report of forty-five cases. J. Thorac. Cardiovasc. Surg. 123, 1125–1131 (2002)
3. Feyter, P.J., Krestin, G.P. (eds.): Computed Tomography of the Coronary Arteries, 2nd edn. Informa Healthcare (2008)
4. Schaap, M., Metz, C., van Walsum, T., van der Giessen, A., Weustink, A., Mollet, N., Bauer, C., Bogunovifa, H., Castro, C., Deng, X., Dikici, E., ODonnell, T., Frenay, M., Friman, O., Hernandez Hoyos, M., Kitslaar, P., Krissian, K., Kuhnel, C., Luengo-Oroz, M., Orkisz, M., Smedby, O., Styner, M., Szymczak, A., Tek, H., Wang, C., Warfield, S., Zambal, S., Zhang, Y., Krestin, G., Niessen, W.: Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms. Med. Image Anal. 13(5), 701–714 (2009)

5. Lesage, D., Angelini, E.D., Funka-Lea, G., Bloch, I.: A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes. Med. Image Anal. 13, 819–845 (2009)
6. Metz, C., Schaap, M., van Walsum, T., van der Giessen, A., Weustink, A., Mollet, N., Krestin, G.P., Niessen, W.: Editorial: 3D segmentation in the clinic: A grand challenge II - coronary artery tracking. In: MICCAI 2008 Workshop Proceedings (2008)
7. Shechter, G., Devernay, F., Quyyumi, A., Coste-Maniere, E., McVeigh, E.: Three-dimensional motion tracking of coronary arteries in biplane cineangiograms. IEEE Trans. Med. Imaging 22(4), 493–603 (2003)
8. Shechter, G., Resar, J.R., McVeigh, E.R.: Displacement and velocity of the coronary arteries: cardiac and respiratory motion. IEEE Trans. Med. Imaging 25, 369–375 (2006)
9. Metz, C., Schaap, M., Klein, S., Neefjes, L., Capuano, E., Schultz, C., van Geuns, R.J., Serruys, P.W., van Walsum, T., Niessen, W.J.: Patient specific 4D coronary models from ECG-gated CTA data for intra-operative dynamic alignment of CTA with X-ray images. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 369–376. Springer, Heidelberg (2009)
10. Beg, F.M., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic ows of diffeomorphisms. International Journal of Computer Vision 61(2), 139–157 (2005)
11. Ceritoglu, C., Oishi, K., Li, X., Chou, M., Younes, L., Albert, M., Lyketsos, C., van Zijl, P., Miller, M., Mori, S.: Multi-contrast large deformation diffeomorphic metric mapping for diffusion tensor imaging. Neuroimage 47(2), 618–627 (2009)
12. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. 12, 26–41 (2008)
13. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage 45(1), S61–S72 (2009)
14. Chillet, D., Jomier, J., Cool, D., Aylward, S.R.: Vascular atlas formation using a vessel-to-image affine registration method. In: Ellis, R.E., Peters, T.M. (eds.) MICCAI 2003. LNCS, vol. 2878, pp. 335–342. Springer, Heidelberg (2003)
15. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J.: Non-rigid registration using free-form deformations: application to breast MR images. IEEE Trans. Med. Imaging 18(8), 712–721 (1999)
16. Manniesing, R., Viergever, M., Niessen, W.: Vessel enhancing diffusion - a scale space representation of vessel structures. Med. Image Anal. 10, 815–825 (2006)
17. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man and Cybernetics 9(1), 62–66 (1979)
18. Frangi, A., Niessen, W., Hoogeveen, R., van Walsum, T., Viergever, M.: Model-based quantitation of 3D magnetic resonance angiographic images. IEEE Trans. Med. Imaging 18(10), 946–956 (1999)
19. Weustink, A.C., Mollet, N.R., Pugliese, F., Meijboom, W.B., Nieman, K., Heijenbrok-Kal, M.H., Flohr, T.G., Neefjes, L.A., Cademartiri, F., de Feyter, P.J., Krestin, G.P.: Optimal electrocardiographic pulsing windows and heart rate: Effect on image quality and radiation exposure at dual-source coronary CT angiography. Radiology 248(3), 792–798 (2008)
20. Jia, J., Tang, C.: Image repairing: Robust image synthesis by adaptative ND tensor voting. In: IEEE CVPR, vol. 1, p. 643 (2003)
21. Risser, L., Plouraboue, F., Descombes, X.: Gap filling of 3-D microvascular networks by tensor voting. IEEE Trans. Med. Imaging 27(5), 674–687 (2008)

A Continuity Equation Based Optical Flow Method for Cardiac Motion Correction in 3D PET Data

Mohammad Dawood[1,2,⋆], Christoph Brune[2], Xiaoyi Jiang[2], Florian Büther[1], Martin Burger[2], Otmar Schober[3], Michael Schäfers[1,3], and Klaus P. Schäfers[1]

[1] European Institute for Molecular Imaging, University of Münster, Mendelstr. 11, 48149 Münster, Germany
dawood@uni-muenster.de
[2] Department of Mathematics and Computer Science, University of Münster, Einsteinstr. 62, 48149 Münster, Germany
[3] Department of Nuclear Medicine, University Hospital Münster Albert-Schweitzer-Str. 33, 48149 Münster, Germay

Abstract. Cardiac Motion artifacts in PET are a well known problem. The heart undergoes two types of motion, the motion due to respiratory displacement and the motion due to cardiac contraction. These movements lead to blurring of data and to inaccuracies in the quantification. In this study a continuity equation based optical flow method is presented and results on 3D PET patient datasets for cardiac motion correction are presented. The method was evaluated with respect to three criteria: correlation between the images, myocardial thickness and the blood pool activity curves. The results showed that the method was successful in motion correcting the data with high precision.

Keywords: Motion correction, Optical Flow, PET, CT, Mass Conservation.

1 Introduction

PET (Positron Emission Tomography) is one method of acquiring metabolic information in patient studies, e.g. to visualize and quantify glucose metabolism in the body. To achieve this, a radioactive substance is injected in the patient body prior to image acquisition. The radioactive isotope decays with time and emits radiation which can be detected in specially built scanners. The distribution of the radioactivity in the body can thus be visualized and gives information on the metabolism. In PET, β^+ radioactive molecules are used for this purpose. These molecules emit positrons which collide with electrons and produce two gamma quanta which fly away from each other in opposite directions. The gamma quanta can now be detected in the scintillation detectors of the PET scanner. Using specialized reconstruction methods the activity distribution can thus be reconstructed [1].

⋆ Corresponding author.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 88–97, 2010.

As this process of image acquisition requires a relatively long period of time, typically several minutes, the motion of the heart due to respiration and due to the cardiac contraction blurs the images. Image blur may cause wrong staging [2], inaccurate localization [4] and wrong quantification [5] of lesions. Thus PET studies have to take this into account.

This problem is compounded further if computed tomography (CT) data is used for attenuation correction, as in the case of modern PET/CT scanners. The CT data represents a snapshot in comparison to the PET images and therefore, the PET data is not always in spatial correspondence with the CT data.

One method of avoiding this problem is to use respiratory and cardiac signals from the patient to divide the PET data into phases with respect to either or both signals [6]. This is called gating. However, gating leads to reduction of the amount of information per phase. To get the same amount of information the image acquisition time has to be proportionally extended or the amount of radioactivity has to be increased. The first option is costly whereas the second exposes the patients to increased radioactivity.

Most recent studies related to this problem estimate the motion on the high resolution and less noisy gated CT images [7],[8],[9]. But this comes at the cost of an increased exposure of the patient to x-rays, which should be avoided where possible.

Two important studies related to the correction for cardiac motion are [10] and [11]. In the first study optical flow is used for estimating the deformations in the images by modeling the myocardium as an elastic membrane. The second study combines the motion estimation of the first study with reconstruction in a single framework. However this study is confined to 2D images and deals with cardiac motion.

1.1 Aim

The aim of this study is to present a new method based on optical flow which can correct the PET images for cardiac motion and is also computationally simple enough to allow reasonable times for motion correction. The method is essentially different from the brightness consistency based optical flow methods [12],[16] as it is based on the continuity equation. This change in the basic model is necessary as brightness consistency is not given in cardiac gated PET data due to the partial volume effect (PVE).

The PVE is a result of the limited resolution of the scanners. All objects smaller than the scanner resolution limit can not be accurately delimited and therefore appear blurred. As the heart muscle contracts and expands during the cardiac cycle its thickness varies. In phases with thicker heart wall, the activity is better resolved and has a higher amplitude as compared to other phases where the activity is spread over a larger area. However, the total amount of the activity remains the same.

The presented method is thus applicable to gated cardiac PET data. It is evaluated on software phantom and real patient data.

2 Intensity and Mass Conserving Optical Flows

Optical flow methods estimate the motion between two image frames. As a voxel with intensity $I(x, y, z, t)$ moves between the two frames, its intensity is assumed to remain constant in intensity conserving optical flow methods. Therefore the following equation holds [16]:

$$I(x, y, z, t) = I(x + \delta x, y + \delta y, z + \delta z, t + \delta t) \tag{1}$$

Here x, y, z are the spatial 3D coordinates and t is the time. Assuming the movement to be small enough and with Taylor expansion we get:

$$\begin{aligned} I_x u + I_y v + I_z w &= -I_t \quad \text{or} \\ \nabla I \cdot \mathbf{u} &= -I_t \end{aligned} \tag{2}$$

with u, v, w for the x,y and z components of the velocity or optical flow $\mathbf{u}$, and I_x,I_y,I_z,I_t for the derivatives of the intensity image I in corresponding directions, respectively time.

To find the optical flow from this equation with three unknowns, additional constraints are required. Smoothness in flow is one such constraint. The famous optical flow algorithm by Horn/Schunck [17] also uses this constraint. The optical flow is thus found using an iterative scheme, whereby an energy functional is minimized. This functional can be given as:

$$f = \min \int ((\nabla I \cdot \mathbf{u} + I_t)^2 + \alpha(|\nabla u|^2 + |\nabla v|^2 + |\nabla w|^2))\mathrm{d}x\mathrm{d}y\mathrm{d}z \tag{3}$$

where larger values of α lead to a smoother flow. The minimization can be achieved by calculating the corresponding Euler-Lagrange equations.

Such methods have been applied to the problem of respiratory motion on 3D PET/CT data successfully [15].

The optical flow estimation presented so far is applicable to data where the intensity of the objects remains constant. However, in some cases this constraint does not hold true. Cardiac PET studies are one such example. As the resolution of the PET scanners is limited, the real radioactive intensity present in an object cannot be accurately located below the limit of resolution resulting in image blur. In case of the heart, the myocardium (the heart muscle) expands and becomes thin in the end-diastolic phase whereas it contracts and becomes thick in the end-systolic phase (see figure 1). Therefore, the intensity seen on PET images of the myocardium varies largely depending upon the heart phase under observation.

In the case of cardiac PET images, another approach is thus required. The here presented method is based upon the continuity equation, more precisely upon the conservation of mass. This law says that the mass in a closed system is conserved. If we substitute activity by mass, the law must still hold, as the total activity in the system remains same from systolic to diastolic phases of the heart. It is only blurred in the diastolic phase. It should be noted that our data is pre-corrected for the time dependent radioactive decay during the reconstruction process so that the decay itself plays no role for our considerations.

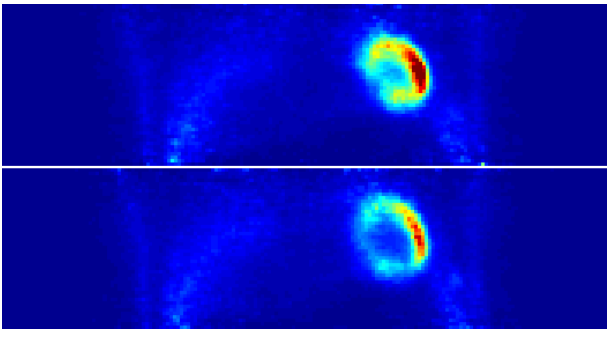


Fig. 1. Two phases from the cardiac cycle of the heart. Above: End-Systole, Below: End-Diastole. A coronal slice from the 3D PET image volume is shown. Images from an FDG study are shown here without attenuation correction.

The continuity equation for mass conservation is given as [13]:

$$\frac{\partial I}{\partial t} + div(I\mathbf{u}) = 0 \tag{4}$$

where I is the intensity value, $\mathbf{u} = (u, v, w)^T$ is the velocity vector i.e. the optical flow. Deviations from this eqution can be penalized by the following functional:

$$\int (\nabla I \cdot \mathbf{u} + I_t + I \cdot div(\mathbf{u}))^2 \mathrm{d}x\mathrm{d}y\mathrm{d}z \tag{5}$$

The derivative in time I_t can be calculated on discrete image volumes by using the difference: $I_2 - I_1$, where I_2 is the floating and I_1 is the target image volume. As with the intensity based optical flow, this is again an under-determined system of equations and therefore a smoothing term can be added to solve it. In the present study we used the same smoothing term as given in equation 3. The resulting optical flow functional is thus:

$$\begin{aligned} f &= \operatorname{argmin}\left[\int_V D^2 dV + \alpha \int_V S dV\right] \\ &\text{with} \\ D &= div(I\mathbf{u}) + I_t, \quad S = |\nabla u|^2 + |\nabla v|^2 + |\nabla w|^2 \end{aligned} \tag{6}$$

The minimization of the equation (6) can be achieved by using the corresponding Euler-Lagrange equations. These are given by:

$$\begin{aligned} 0 &= D_x I + \alpha \Delta u \\ 0 &= D_y I + \alpha \Delta v \\ 0 &= D_z I + \alpha \Delta w \end{aligned} \tag{7}$$

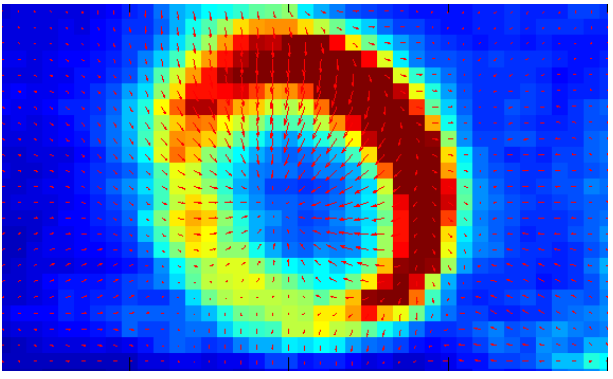


Fig. 2. The optical flow calculated with the proposed method. A coronal slice is shown with superimposed vectors. Only two components of the flow are shown.

where D_x, D_y, D_z are the first derivatives of D in the corresponding directions. The weighting parameter α was set to about 0.05 based upon our previous experience with the data.

Once the optical flow is found (see Figure 2), the images have to be transformed to get the motion corrected data. The equation (4) can be used for this purpose. As the time derivative I_t was calculated as $I_2 - I_1$ and the flow $\mathbf{u}$ is now assumed to be known, the transformed image can be calculated as:

$$I_{2mc} = I_2 + div(I_2\mathbf{u}) \tag{8}$$

3 Software Phantom and Patient Data

To validate the methods, software phantom data was used. The NCAT software phantom by segars et al [14] provides a widely used tool for emission tomographic data simulation. The phantom data were produced for a cardiac cycle with 10 phases. The first gate was set to be end-diastolic and thus the end-systolic phase was gate number 5. The voxel size was 3.125 mm in each direction. Default parameters for the size and activity were used. The data was noise free and contained only minimal partial volume effects.

Fourteen patients with known coronary artery disease were included in this study. Patients were routinely referred to the ^{18}FDG PET scan for evaluation of myocardial viability prior to revascularization. A listmode dataset was acquired for 20 minutes, 1 hour post injection of ^{18}FDG (4 MBq/kg). To enhance FDG uptake in the heart, the patients underwent a hyperinsulinemic euglycemic clamp technique prior to and during the scan [18]. All patients received β-blockers to slow down and stabilize the heart rate during CT examination.

The Siemens Biograph Sensation 16 PET/CT scanner (Siemens Medical Solution) with a dedicated listmode research package was used in these studies. This PET scanner has a spatial resolution of around 6 mm [19]. The cardiac signal

for gating was acquired during the PET acquisition. The listmode file contains the coincidences along with the time of occurrence. This information together with the ECG information was used to sort the data into 10 cardiac phases. The data was then reconstructed without attenuation correction with the help of an expectation maximization algorithm [3]. In patient datasets, the end-diastolic phase was gate 3 and the end-systolic phase was gate number 9.

4 Results and Discussion

The performance of the proposed method was estimated with help of three criteria. These are 1) the correlation coefficient, 2) the myocardial thickness and 3) the left ventricular blood pool activity. The results of these experiments are given below in the corresponding subsections.

In all studies the diastolic phase was used as the target gate. It should be remembered that the cardiac cycle does not follow a linear pattern. This means that some phases are very close to each other and there are large differences among others.

4.1 Correlation Coefficient

The spatial correlation between the original and the motion corrected volumes was calculated between the target phase (end-diastolic) and all other phases before and after motion correction. To discard the influence of the stationary voxels, such as out of body pixels, a 40x40x40 voxels large volume of interest (VOI) was selected around the heart. The correlation coefficient was then calculated as:

$$cc = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \bar{A})^2)(\sum_m \sum_n (B_{mn} - \bar{B})^2)}} \tag{9}$$

where $\bar{A}$ and $\bar{B}$ are the means intensities of the respective volume.

The results for the software phantom are given in Table 1. The average correlation after motion correction is 99.86 as compared to 88.51 before the same. Besides high correlation, it is to be noted that the variation among the uncorrected data is far greater (min 80.25, std: 2.67) than after motion correction (min 99.81, std: 0.23). This shows that the algorithm effectively corrected the motion for all phases.

Table 1. Results of the correlation analysis on software phantom dataset

Phase	1	2	3	4	5	6	7	8	9	10	Avg
Before MC	100.0	94.53	92.99	86.65	81.73	80.25	80.75	82.22	90.88	95.08	88.51
After MC	100.0	99.87	99.87	99.85	99.83	99.81	99.82	99.83	99.87	99.87	99.86

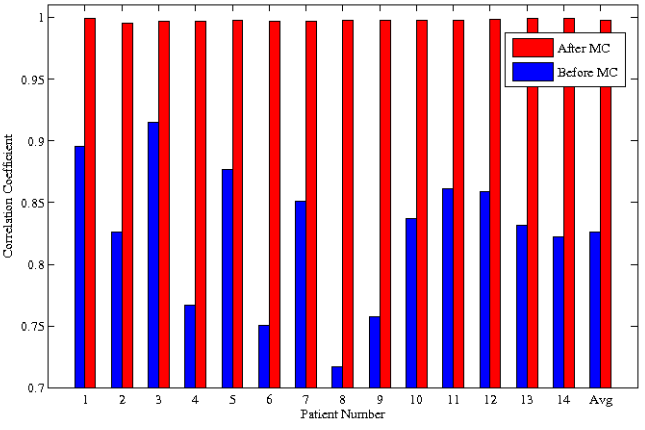


Fig. 3. Results of the correlation analysis on systolic phase of all patient datasets

An overview of results on patient data can be seen in Figure 3. For better readability the results are given only for the correlation of the end-systolic with the target phase. Again, a high mean of 99.76 for all patients (min 99.54, std: 0.11) as compared to the original data (min 71.70, std: 5.84) was achieved.

4.2 Myocardial Thickness

The thickness of the left ventricular wall increases from the end-diastolic to the end-systolic phase to pump blood into the arteries. As this variation follows the cardiac cycle, the wall thickness for all phases should correspond to that of the target phase after motion correction.

To calculate this measure, line profiles were taken across the left ventricular wall after manual reangulation of the heart. The distance between the ascending and descending flanks of the profile curve was taken as the width of the myocardial wall. For this the line profile was fitted with a gaussian curve and the FWHM of the fittgin function was used. To reduce the influence of noise, three consecutive slices were selected in the mid ventricular area and the average wall thickness calculated. As opposed to the correlation coefficient, this measure is more localized in nature.

The results on phantom data are given in Table 2 and show that the myocardial thickness is consistently (within 0.1 mm) similar to that of the target phase

Table 2. Myocardial thickness [mm] on software phantom dataset

Phase	1	2	3	4	5	6	7	8	9	10
Before MC	10.0	10.1	10.7	12.6	13.6	13.0	12.3	11.3	10.7	10.2
After MC	10.0	9.9	9.9	10.0	9.9	9.9	9.9	9.9	9.9	9.9

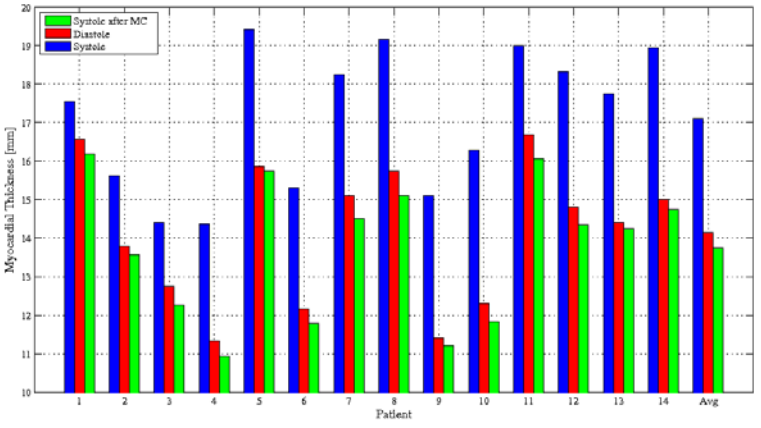


Fig. 4. Results of the myocardial thickness analysis on all patient datasets

(to be recalled: voxel size was 3.125 mm^3). The standard deviation was reduced from 1.32 mm to 0.04 mm.

The results for patient data are given in Figure 4. The wall thickness was 17 mm on average in the end-systolic phase, which compared to 14 mm on average for the end-diastolic phase. After motion correction the new transformed end-systolic phase also showed a wall thickness of 14 mm on average. The sum of squared differences between the wall thickness in end-systolic and end-diastolic phases was reduced from 133.9 to 2.5 after motion correction. The wall thickness appears larger than usual as attenuation correction was not performed.

4.3 Mean Activity in Blood Pool

The third criterion to assess the performance was the mean activity in the blood pool in the left ventricle. Due to the PVE, activity radiates from the myocardium into the blood pool. In the end-systolic phase the blood pool is small, the ventricular walls closer to each other and accordingly there is a greater blurring effect as compared to the end-diastolic phase, when the blood pool is larger and the walls farther apart. After motion correction the blood activity values should be similar for all phases. For this analysis a 6x6x6 voxels large VOI was selected inside the left ventricle manually for each data set and the mean activity calculated in this VOI.

The results given in Figure 5 show that the activity in the blood pool becomes relatively independent of the cardiac phase after the proposed algorithm is applied. In the case of patient 10, the patient with the highest myocardial uptake on the Figure, the standard deviation of mean VOI activity among all cardiac phases was reduced from 409.5 to 26.5. The radiation into the blood increases with higher uptake in the myocardium. Consequently, the patients with high myocardial uptake show the largest variance among blood pool activities of the systolic and diastolic phases.

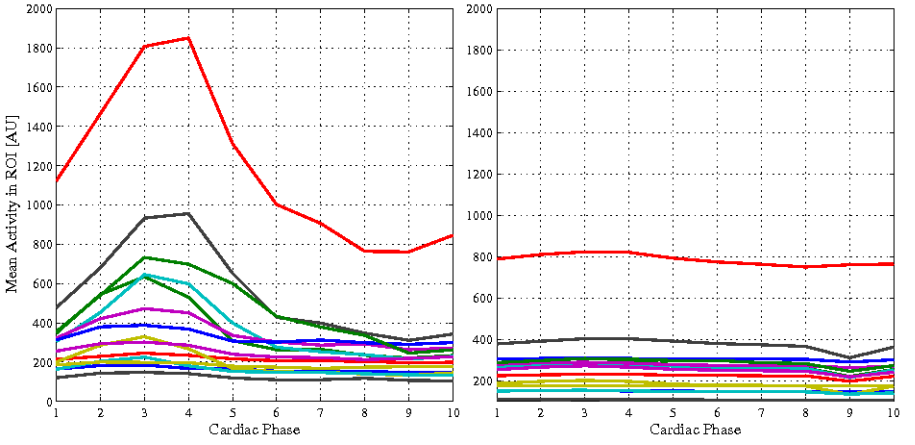


Fig. 5. Results of the mean activity in blood pool analysis among all patient datasets.

5 Conclusions

A continuity equation based optical flow method for cardiac PET data motion correction is presented. The method is able to correct the data despite partial volume effect. It was validated on patient and software phantom data. The results showed that the cardiac motion was corrected precisely.

Acknowledgments. This project was funded by the Deutsche Forschungsgemeinschaft SFB 656 (projects B2 & B3) and by the Deutsche Telekom Foundation. We thank our collegues Gigengack and Ruthotto for important discussions.

References

1. Vandenberghe, S., D'Asseler, Y., Van de Walle, R., Kauppinen, T., Koole, M., Bouwens, L., Van Laere, K., Lemahieu, I., Dierckx, R.A.: Iterative reconstruction algorithms in nuclear medicine. Comput. Med. Imaging Graph. 25(2), 105–111 (2001)
2. Erdi, Y.E., Nehmeh, S.A., Pan, T., Pevsner, A., Rosenzweig, K.E., Mageras, G., Yorke, E.D., Schoder, H., Hsiao, W., Squire, O.D., Vernon, P., Ashman, J.B., Mostafavi, H., Larson, S.M., Humm, J.L.: The CT motion quantitation of lung lesions and its impact on PET-measured SUVs. J. Nucl. Med. 45(8), 1287–1292 (2004)
3. Shepp, L.A., Vardi, Y.: Maximum Likelihood Reconstruction for Emission Tomography. IEEE Trans. Med. Imag. 1(2), 113–122 (1982)
4. Osman, M.M., Cohade, C., Nakamoto, Y., Marshall, L.T., Leal, J.P., Wahl, R.L.: Clinically significant inaccurate localization of lesions with PET/CT: frequency in 300 patients. J. Nucl. Med. 44(2), 240–243 (2003)
5. Nakamoto, Y., Chin, B.B., Cohade, C., Osman, M., Tatsumi, M., Wahl, R.L.: PET/CT: artifacts caused by bowel motion. Nucl. Med. Commun. 25(3), 221–225 (2004)

6. Dawood, M., Büther, F., Lang, N., Schober, O., Schäfers, K.P.: Respiratory gating in positron emission tomography: A quantitative comparision of different gating schemes. Medical Physics 34, 3067–3076 (2007)
7. Qiao, F., Pan, T., Clark, J.W., Mawlawi, O.R.: A motion-incorporated reconstruction method for gated PET studies. Phys. Med. Biol. 51(15), 3769–3783 (2006)
8. Mair, B.A., Gilland, D.R., Sun, J.: Estimation of images and nonrigid deformations in gated emission CT. IEEE Trans. Med. Imaging 25(9), 1130–1144 (2006)
9. Lamare, F., Carbayo, M.J.L., Cresson, T., Kontaxakis, G., Santos, A., Rest, C.C.L., Reader, A.J., Visvikis, D.: List-mode-based reconstruction for respiratory motion correction in PET using non-rigid body transformations. Phys. Med. Biol. 52, 5187–5204 (2007)
10. Klein, G.J., Reutter, B.W., Huesman, R.H.: Non-rigid summing of gated PET via optical flow. IEEE Transactions on Nuclear Science 44(4), 1509–1512 (1997)
11. Gilland, D.R., Mair, B.A., Bowsher, J.E., Jaszczak, R.J.: Simultaneous reconstruction and motion estimation for gated cardiac ECT. IEEE Transactions on Nuclear Science 49(5), 2344–2349 (2002)
12. Dawood, M., Büther, F., Jiang, X., Schäfers, K.P.: Respiratory Motion Correction in 3D PET Data with Advanced Optical Flow Algorithms. IEEE Trans. Med. Imaging 27(8), 1164–1175 (2008)
13. Corpetti, T., Heitz, D., Arroyo, G., Mémin, E., Santa-Cruz, A.: Fluid experimental flow estimation based on an optical-flow scheme. Experiments in Fluids 40, 80–97 (2006)
14. Segars, W.P.: Development and application of the new dynamic NURBS-based cardiac-torso (NCAT) phantom, in Biomedical Engineering. Dissertation, University of North Carolina (2001)
15. Dawood, M., Lang, N., Jiang, X., Schäfers, K.P.: Lung motion correction on respiratory gated 3-D PET/CT images. IEEE Trans. Med. Imaging 25(4), 476–485 (2006)
16. Bruhn, A., Weickert, J., Schnörr, C.: Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. International Journal of Computer Vision 61(3), 211–231 (2005)
17. Horn, B., Schunck, B.: Determining optical flow. Artificial Intelligence 17, 185–203 (1981)
18. Vitale, G.D., de Kemp, R.A., Ruddy, T.D., Williams, K., Beanlands, R.S.: Myocardial glucose utilization and optimization of (18)F-FDG PET imaging in patients with non-insulin-dependent diabetes mellitus, coronary artery disease, and left ventricular dysfunction. J. Nucl. Med. 42(12), 1730–1736 (2001)
19. Erdi, Y.E., Nehmeh, S.A., Mulnix, T., Humm, J.L., Watson, C.C.: PET performance measurements for an LSO-based combined PET/CT scanner using the national electrical manufacturers association nu 2-2001 standard. J. Nucl. Med. 45(5), 813–821 (2004)

Simultaneous Reconstruction of 4-D Myocardial Motion from Both Tagged and Untagged MR Images Using Nonrigid Image Registration

Wenzhe Shi, Maria Murgasova, Philip Edwards, and Daniel Rueckert

Visual Information Processing Group
Department of Computing
Imperial College London

Abstract. Tagged magnetic resonance (MR) imaging is unique in its ability to noninvasively image the motion and deformation of the heart. However, it is difficult to identify and quantify structures of interest in the cardiac anatomy since the tags obscure the anatomy. In this paper, we present a novel and fully automated technique based on nonrigid image registration for the analysis of myocardial motion using both tagged and untagged MR images. The novel aspect of our technique is its simultaneous usage of complementary information from both tagged and untagged images. No manual intervention is required to obtain the segmentation of the end-diastolic images. To estimate the motion within the myocardium, we register a sequence of images taken during systole to a set of reference images taken at end-diastole, maximizing a spatial weighted similarity measure between the images. We use short-axis and long-axis images of the heart as well as tagged and untagged images to estimate a fully four-dimensional motion field within the myocardium. We have evaluated the proposed approach on 8 patients both in terms of robustness, accuracy and consistency of the motion tracking. The proposed method is significantly more consistent than motion tracking on tagged MR images only.

1 Introduction

The ultimate objective of cardiac image analysis is to provide useful and efficient tools for the diagnosis and treatment of patients with cardiovascular diseases. An increasing amount of attention has been focussed on the estimation of local deformation parameters, such as strain. The analysis of such parameters can help to better understand diseases such as cardiomyopathy and ischemia and can lead to improved methods for the treatment of patients with cardiovascular diseases.

Myocardial tissue in the body can be labeled by altering its magnetization properties which remain persistent even in the presence of motion. MR tagging was first proposed by [17] as a means of non-invasively introducing markers within the myocardium of the left ventricle. Using this technology, noninvasive markers can be introduced directly into the tissue being imaged during the image

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 98–107, 2010.

acquisition process. By tracking the motion and deformation of the tag patterns, the motion of the myocardium can be reconstructed by cardiac motion analysis algorithms [9,16,10] including nonrigid image registration [2], [7], [1], [3]. The cardiac motion is reconstructed by registering a sequence of images taken during the contraction of the heart to a reference image taken at the start of the cardiac cycle.

A common difficulty in model-based and registration-based motion tracking is to address the tag fading. Tagged MR images are analyzed in two steps to deal with this difficulty. The first step requires the identification and tracking of tags in the MR images, which involves the segmentation of the LV wall by identifying its outer and inner contours. The second step involves the reconstruction of 3D cardiac motion and the computation of motion-related variables, such as strain, displacement field, and torsion. It is either difficult or time consuming to automate the first step on tagged MR images. HARP based motion tracking on the other hand, cannot construct real 3D motion since it is inherently 2D. Although [10] reconstructed 3D motion based on the results of 2D in-plane motions, it is more desirable to extract the 3D motion from images directly and consistently.

In this paper we focus on motion tracking using both tagged and untagged MR images simultaneously. An advantage of untagged MR images is that the cardiac anatomy and in particular the myocardium is clearly visible and can be identified using state-of-the-art image segmentation algorithms [6,4,5]. In additional, the radial motion of the myocardium can be tracked easily in untagged MR images since the epi- and endocardial surfaces are clearly visible. A disadvantage of untagged MR images is that circumferential and longitudinal motion cannot be accurately quantified as there are few landmarks inside the myocardium which can be reliably tracked. On the other hand, tagged MR images allows easy tracking of both longitudinal circumferential and radial motion. However in tagged MR images it is difficult to identify and quantify the cardiac anatomy since the tags obscure the anatomy. Although tag removal technologies exists [12], the quality of the resulting images is not as good as conventional untagged MR images. Moreover, it is especially difficult to remove grid-like tag patterns, which are often used as they help to reduce acquisition time and registration complexity. The lack of visible anatomy in the tagged MR images can cause problems during the motion tracking as it is difficult to differentiate between tissue and fluids (e.g. the blood pool). Nevertheless, both types of MR images sequences provide complementary information that can be exploited.

One of the advantages of the combined usage of tagged and untagged MR images is that both anatomical and tagging image information can be used at the same time. Another advantage is that the anatomical information in the untagged MR images enables a wide range of potential automatic segmentation technologies to be used in the registration process. This eliminates the need to manually segment the myocardium.

In this paper we developed a combined registration and motion tracking algorithm using both tagged and untagged images. For this purpose we have extended a registration algorithm which has been previously used successfully for motion

tracking [2] . Before registration we use a Haar-feature based object detection algorithm [15] [11] to detect a region of interest containing the left ventricle. Because the motion of the heart is 3D, both short- (SA) and long-axis (LA) images of the heart are used. A spatially-varying, weighted similarity measure is used for the image registration. This similarity measure combines information from both LA and SA images as well as from tagged and untagged images. The weighting between the different images is spatially varying and depends on the intensity gradient and probabilistic segmentation of the untagged image: At the epicardial and endocardial boundaries (indicated by high intensity gradients in the untagged images), the weighting favours the untagged images. Inside the myocardium (indicated by the EM segmentation of the untagged images) the weighting favours the tagged images.

The next section describes the proposed approach and details of the combined registration algorithm. Section 3 describes the evaluation of the proposed technique in terms of accuracy, robustness and consistency using data from 8 patients. Finally, section 4 includes a discussion of the results and future work.

2 Combined Registration

During the systolic phase, the left ventricle undergoes a number of different types of deformations including: twisting motion, radial motion and longitudinal motion. Since the imaging planes defined in the MR scanner are stationary with respect to the coordinate system of the scanner, the motion of the heart is not confined to a single plane during the cardiac cycle. Thus, to fully reconstruct the deformation field within the myocardium, we need to acquire multi-slice SA and LA images of the LV.

Consider a material point in the myocardium at a position $\mathbf{p} = (x, y, z)^T$ at time $t_0 = 0$ that moves to another position $\mathbf{p}' = (x', y', z')^T$ at time $t_i = i\Delta t$ where Δt is the time interval between two consecutive frames and i corresponds to the frame number. The problem is to find the transformation $\mathbf{T}$ for all time frames i such that:

$$\mathbf{T}(\mathbf{p}, t_i) = \mathbf{p}' \tag{1}$$

We represent $\mathbf{T}$ using a series of free-form deformations [14] as described in [2]. An overview of the tracking algorithm is given in the following section.

2.1 Overview

The estimation of the deformation field $\mathbf{T}$ proceeds in a sequence of registration steps. We first map all images into the scanner coordinate system using the position and orientation information obtained from the DICOM headers of images. After that, we detect the position of the heart and bounding box in the untagged images using an object detector [15]. Within the bounding box, we use an EM-based segmentation algorithm to classify each voxel according to its tissue type. In addition, a gradient detector is used to identify the epicardial and endocardial contours. The information from both EM segmentation and gradient-detection is

combined into spatially varying weight function which moderates the influence of the tagged and untagged image information. We then register the images taken at time t_1 to the reference image t_0 and obtain a transformation representing the motion of the myocardium at time t_1. We use the resulting transformation as input and continue the process until all the volumes in the sequence are registered to the first frame[2].The algorithm allows us to relate any point in the myocardium at time $t = 0$ to its corresponding position throughout the sequence.

2.2 Detection and Segmentation of the Heart

We use a modified Viola-Jones object detection framework [15] [11] to detect the interest region from the untagged images in the world coordinate.

We use an atlas-based EM segmentation algorithm [5] [8] to segment the MR images into three different classes – background, blood pool and myocardium. An affine registration followed by a non-rigid registration [13] is performed between untagged MR images and the atlas.

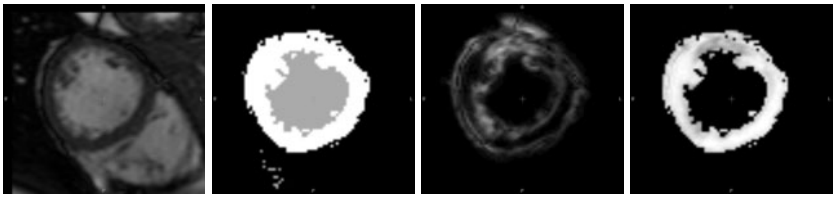


Fig. 1. Four images are from left to right: Untagged MR image, segmentation result, weight map for untagged MR image and weight map for tagged MR image

2.3 Weighted Similarity Measure

To exploit the complementary nature of the tagged and untagged MR images we have developed a spatially varying weight function that accounts for the different types of information available: Tagged images characterize well the motion inside the myocardium while untagged images characterize well the motion at the epi- and endocardial borders of the myocardium. Outside the myocardium, e.g. inside the blood pool or in the lungs, neither images contain any useful information for cardiac motion tracking. Weights are generated based on target image of the registration in our case which is the end diastole frame.

Let L denote the EM-based segmentation of the untagged image I. This segmentation assigns a label $\Lambda = \{L_{bg}, L_{myo}, L_{blood}\}$ to every voxel. Furthermore assume that ∇I_σ denotes the gradient of image I after convolution with a Gaussian kernel G with standard deviation σ. The weights for the untagged image are defined as

$$W^u(\mathbf{p}) = \frac{||\nabla I_\sigma(\mathbf{p})||||\nabla P_\sigma(\mathbf{p}, L_{myo})||}{\max||\nabla I_\sigma(\mathbf{p})||\max||\nabla P_\sigma(\mathbf{p}, L_{myo})||}$$

where $||\nabla I_\sigma(\mathbf{p})||$ is the gradient of intensity at location p and $||\nabla P_\sigma(\mathbf{p}, L_{myo})||$ is the gradient of myocardium probability at location p. The weights for the tagged image are defined as

$$W^t(\mathbf{p}) = \begin{cases} 1 - W^u(\mathbf{p}) & \text{if } L(\mathbf{p}) = L_{myo} \\ 0 & \text{otherwise} \end{cases}$$

An example of the resulting weight maps is shown in Figure 1. Given a weight map, we define the similarity between two images I_A, I_B as the weighted normalized cross-correlation between the image intensities:

$$C(I_A; I_B; W, \mathbf{T}) = \frac{\sum W(\mathbf{p})(I_A(\mathbf{p}) - \mu_A)(I_B(\mathbf{T}(\mathbf{p})) - \mu_B)}{\sqrt{\sum W^2(\mathbf{p})(I_A(\mathbf{p}) - \mu_A)^2 (I_B(\mathbf{T}(\mathbf{p})) - \mu_B)^2}} \tag{2}$$

Here μ_A and μ_B denote the average intensities in image I_A and I_B respectively. For simultaneous registration of the tagged and untagged images, the correlation across tagged images and the correlation across untagged images are combined into a similarity measure as:

$$S = \sum_{s \in LA,SA} |\Omega_s| \left[C(I^{t,s}_{t_0}, I^{t,s}_{t_i}, W^{t,s}, \mathbf{T}) + C(I^{u,s}_{t_0}, I^{u,s}_{t_i}, W^{u,s}, \mathbf{T}) \right] / \sum_{s \in LA,SA} |\Omega_s|$$

Here $|\Omega_{SA}|$ and $|\Omega_{LA}|$ denotes the number of voxels in the short- and long-axis images. Note, that the similarity takes into account that the short- and long-axis images have usually a different number of voxels and the correlation must be weighted accordingly. Using the above similarity measure, the registration relies on the untagged image at those voxels where there is a high gradient value indicating possible presence of an edge outside the myocardium and on tagged images where inside the myocardium.

3 Evaluation

In our experiments we have used images from 8 patients. For each patient four different image sequences were acquired during the same scanning session which correspond to short- and long-axis with and without tagging. All images were acquired on a Siemens Sonata 1.5 T scanner consisting of 10 SA and 3 LA slices covering the whole of the LV. A cine breath-hold sequence with a SPAMM tag pattern was used with imaging being done at end expiration. The image voxel sizes were $1.45 \times 1.45 \times 8$mm , with the distance between slices being 10mm .14 frames for all sequences were acquired during the cardiac cycle. The imaging parameters used a repetition time of 40 ms, an echo time of 4 ms, and a 15^o flip angle. An example data set is shown in Figure 2 together with the reconstructed motion fields.

The evaluation of algorithms for cardiac motion is often difficult because of the lack of a gold standard for motion tracking. Instead, one has to compare the

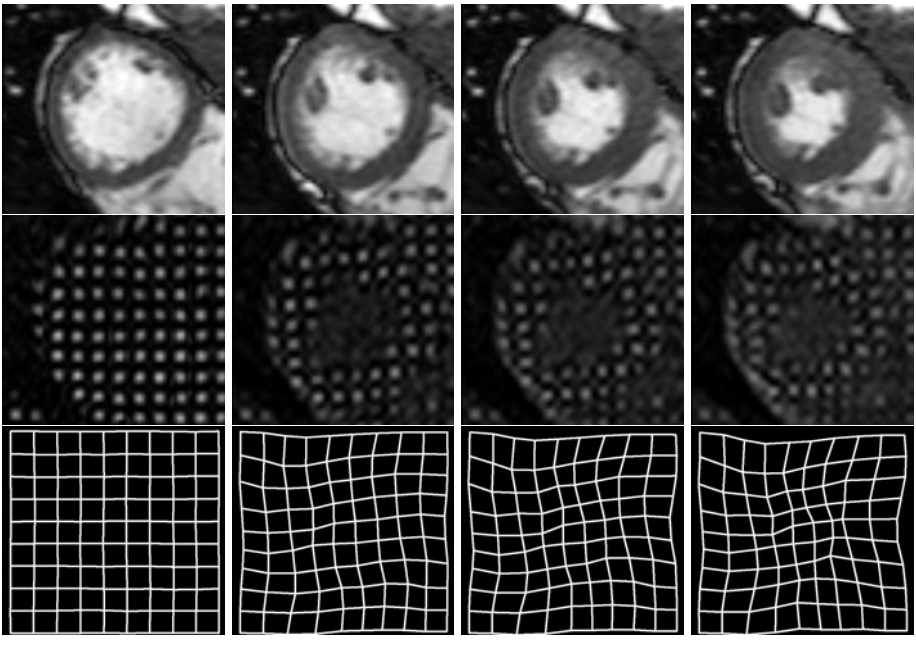


Fig. 2. Untagged (top row), tagged (middle row) MR images and the extracted motion fields (bottom row) from one volunteer

motion tracking to other established techniques. In our application, the problem is compounded by the fact that any comparison to techniques using tagged or untagged images only will be biased. For each data set we have manually tracked 32 landmarks in-plane in the tagged short-axis images and 9 additional landmarks in the tagged long-axis images. During the manual landmark tracking, the position of the landmarks is automatically refined using a center-of-gravity operator. This allows landmark tracking with sub-voxel accuracy. To assess the quality of the registration we compare the position of the manually tracked landmarks with the landmark position predicted by the nonrigid registration. The RMS distance between the landmarks is shown in Figure 3.

The results indicate that the registration using tagged and untagged images with a mean of 1.85mm performs much better than using only untagged images with a mean of 3.44mm. However, the registration using tagged images only performs slightly better with a mean of 1.67mm. This is expected since the manual landmarks have been identified in the tagged images. The evaluation is therefore inherently biased towards tagged images. We have also evaluated the relative error which is the $\sum_{i=1}^{n} \frac{\epsilon}{d}/n$ where the d is the magnitude of the displacement estimated from the transformation and ϵ is the distance between the transformed original point and the manually tracked evaluation point. The relative error of our combined method is very close to the tagged only method on average and better in many cases figure 4. For frame 7 the relative error of

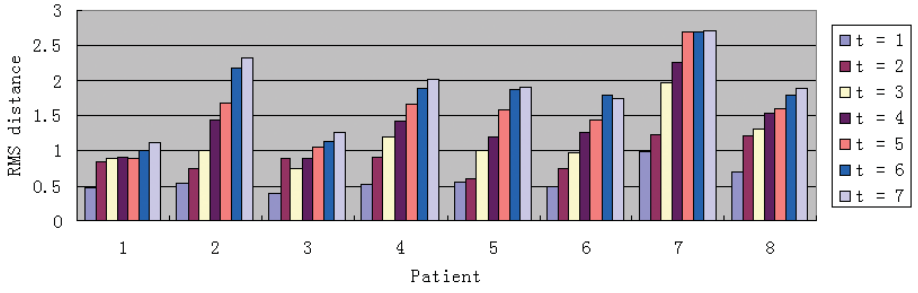


Fig. 3. This figure shows the tracking error when comparing the result of manual tag tracking with registration-based motion tracking. Results are shown for each time frame in 8 datasets.

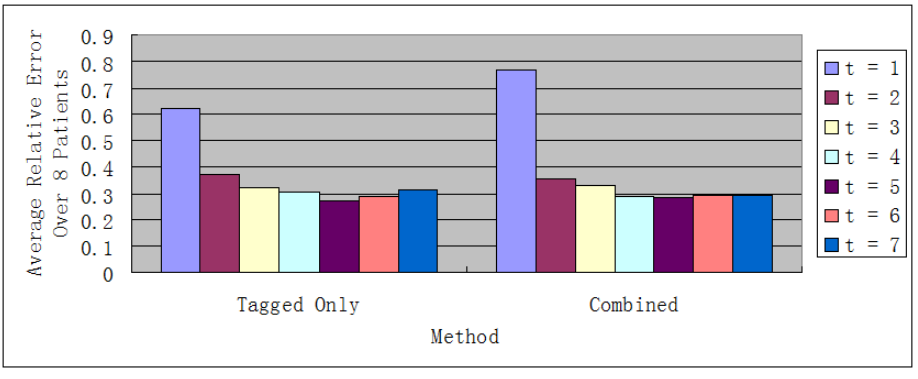


Fig. 4. Relative error when comparing the result of manual tag tracking with registration-based motion tracking normalized by the deformation's distance. Average results are shown for each time frame over 8 datasets of two different methods

the combined method is 0.29 compare to 0.31 for the tagged only method and 0.64 for the untagged only method. The combined method has similar accuracy to the tagged only method.

A limitation of the tagged only methods is that, if there is a sufficiently large motion between two time frames, the motion tracking algorithms may become confused and report no motion at all. This can happen if the number of slices acquired and the temporal resolution of the images is not sufficient to capture the deformation of the myocardium on tags. Thus an alternative strategy for evaluating the performance of the algorithm is by measuring the consistency of the motion tracking. This provides a measure of the robustness and can be defined as the RMS distance between $\mathbf{T}(\mathbf{p}, t_i)$ and $\mathbf{T}'(\mathbf{p}, t_i)$ as shown in Figure 5. The result in Figure 6 show that the consistency of the motion tracking is 8 times better if tagged and untagged images are used simultaneously compared to the use of tagged images only. Figure 4 shows registration done consecutively versus skip one frame. If the registration is consistent then the distance between

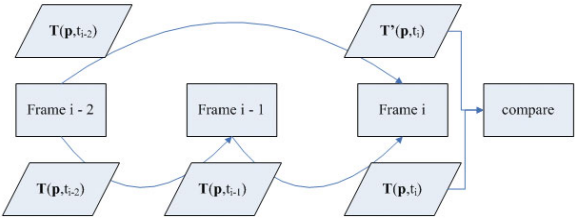


Fig. 5. Definition of consistency in the context of motion tracking. $\mathbf{T}(\mathbf{p}, t_i)$ is generated by optimizing $\mathbf{T}(\mathbf{p}, t_{i-1})$ with image frame 0 and frame t_i while $\mathbf{T}'(\mathbf{p}, t_i)$ is generated by optimizing $\mathbf{T}(\mathbf{p}, t_{i-2})$ with image frame 0 and frame $t - i$.

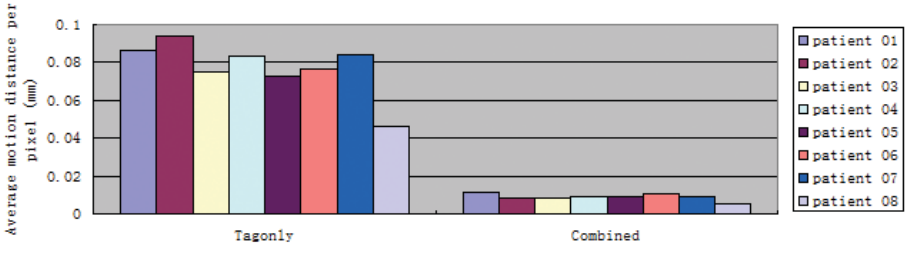


Fig. 6. Consistent measure from tagged-only registration and combined registration, the average distance of motion of voxels in every pair $\mathbf{T}(\mathbf{p}, t_i)$ and $\mathbf{T}'(\mathbf{p}, t_i)$ is plotted in this figure with respect of the patients. Combined registration provided 8 times better consistency than the tagged only registration.

$\mathbf{T}(\mathbf{p}, t_i)$ and $\mathbf{T}'(\mathbf{p}, t_i)$ should be small in order to be able to recover from coarse frames, noise and lose of information. A small value indicates higher robustness. This could be achieved by a consistent but inaccurate registration. However the combined evidence of the accuracy in the landmark tracking and the consistency suggests that our method works well.

4 Conclusion

We have presented a novel method for cardiac motion tracking using both tagged and untagged image sequences from short and long axis simultaneously. The key advantage of the proposed method is the simultaneous exploitation of complementary information contained in the tagged and untagged images. Evaluation shows that there is a significant improvement of consistency compare to tagged only methods. We think it due to improved ability to track larger deformation with spatial weight emphasizes on myocardium edges in untagged images.

The advantage of the presented approach is to reconstruct real 3D motion directly from the MR images. We provide a fully automated registration scheme, which is robust to the tag fading and provides a more consistent and presumably more accurate motion tracking.

References

1. Bistoquet, A., Oshinski, J., Škrinjar, O.: Myocardial deformation recovery from cine MRI using a nearly incompressible biventricular model. Medical Image Analysis 12(1), 69–85 (2008)
2. Chandrashekara, R., Mohiaddin, R.H., Rueckert, D.: Analysis of 3-D myocardial motion in tagged MR images using nonrigid image registration. IEEE Transactions on Medical Imaging 23(10), 1245–1250 (2004)
3. Ledesma-Carbayo, M.J., Kybic, J., Desco, M., Santos, A., Suhling, M., Hunziker, P., Unser, M.: Spatio-temporal nonrigid registration for ultrasound cardiac motion estimation. IEEE Transactions on Medical Imaging 24(9), 1113–1126 (2005)
4. Lingrand, D., Charnoz, A., Koulibaly, P.M., Darcourt, J., Montagnat, J.: Toward accurate segmentation of the LV myocardium and chamber for volumes estimation in gated SPECT sequences. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3024, pp. 267–278. Springer, Heidelberg (2004)
5. Lorenzo-Valdés, M., Sanchez-Ortiz, G.I., Elkington, A.G., Mohiaddin, R.H., Rueckert, D.: Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm. Medical Image Analysis 8(3), 255–265 (2004)
6. Lynch, M., Ghita, O., Whelan, P.F.: Left-ventricle myocardium segmentation using a coupled level-set with a priori knowledge. Computerized Medical Imaging and Graphics 30(4), 255–262 (2006)
7. Mansi, T., Peyrat, J.M., Sermesant, M., Delingette, H., Blanc, J., Boudjemline, Y., Ayache, N.: Physically-Constrained Diffeomorphic Demons for the Estimation of 3D Myocardium Strain from Cine-MRI. In: Ayache, N., Delingette, H., Sermesant, M. (eds.) FIMH 2009. LNCS, vol. 5528, pp. 201–210. Springer, Heidelberg (2009)
8. Murgasova, M., Rueckert, D., Edwards, D., Hajnal, J.: Robust segmentation of brain structures in MRI. In: Proceedings of the Sixth IEEE International Conference on Symposium on Biomedical Imaging: From Nano to Macro, pp. 17–20. IEEE Press, Los Alamitos (2009)
9. Osman, N.F., Kerwin, W.S., McVeigh, E.R., Prince, J.L.: Cardiac motion tracking using CINE harmonic phase (HARP) magnetic resonance imaging. Magnetic Resonance in Medicine 42(6), 1048 (1999)
10. Pan, L., Prince, J.L., Lima, J.A.C., Osman, N.F.: Fast tracking of cardiac motion using 3D-HARP. IEEE Transactions on Biomedical Engineering 52(8), 1425–1435 (2005)
11. Pavani, S.K., Delgado, D., Frangi, A.F.: Haar-like features with optimally weighted rectangles for rapid object detection. Pattern Recognition 43(1), 160–172 (2009)
12. Qian, Z., Huang, R., Metaxas, D., Axel, L.: A novel tag removal technique for tagged cardiac MRI and its applications. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Citeseer, pp. 364–367 (2007)
13. Rueckert, D., Frangi, A.F., Schnabel, J.A.: Automatic construction of 3D statistical deformation models using non-rigid registration. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 77–84. Springer, Heidelberg (2001)
14. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Nonrigid registration using free-form deformations: application to breast MR images. IEEE Transactions on Medical Imaging 18(8), 712–721 (1999)

15. Viola, P., Jones, M.: Robust real-time object detection. International Journal of Computer Vision 57(2), 137–154 (2002)
16. Young, A.A.: Model tags: direct three-dimensional tracking of heart wall motion from tagged magnetic resonance images. Medical Image Analysis 3(4), 361–372 (1999)
17. Zerhouni, E.A., Parish, D.M., Rogers, W.J., Yang, A., Shapiro, E.P.: Human heart: tagging with MR imaging–a method for noninvasive assessment of myocardial motion. Radiology 169(1), 59 (1988)

Cortical Sulcal Bank Segmentation via Geometric Similarity Based Graph Partition

Gang Li[1], Lei Guo[1], Tuo Zhang[1], Jingxin Nie[1], and Tianming Liu[2]

[1] School of Automation, Northwestern Polytechnical University, Xian, China
[2] Department of Computer Science, The University of Georgia, Athens, GA, USA

Abstract. Development of imaging predictor of brain cytoarchitecture has been of significant interest in neuroimaging. Cortical geometric characteristic has been shown to be a good predictor of cortical cytoarchitecture. In this paper, a novel method based on sulcal geometric features is proposed for the extraction of sulcal banks, which are the cortical regions bounded by sulcal fundi and adjacent gyral crest lines. Given parcellated sulcal basins, we apply two graph partition techniques including the normalized cuts and the graph cuts to partition a sulcal basin into two opposing sulcal banks. Particularly, we designed novel geometric similarity metrics and cost functions to adopt these two graph partition algorithms specifically for our applications. As a test bed application, we applied this method to extract the anterior and posterior sulcal banks of the central sulci on over 400 cortical surfaces and achieved promising results. In addition, we applied this method to study the cortical thickness of central sulci and found that the cortical thickness of the anterior sulcal bank is significantly thicker than that of the posterior sulcal bank, suggesting that the segmented sulcal banks can differentiate cortical thickness layout, which is believed to be an indicator of brain cytoarchitecture. Finally, the asymmetry and longitudinal changes of cortical thickness are analyzed using the OASIS database and reasonable results are obtained.

Keywords: sulcal bank, sulcal basin, cortical thickness, graph partition.

1 Introduction

Geometric folding characteristic has been shown to be a good predictor of brain cytoarchitecture [1]. For example, many different cytoarchitectonic regions are separated by sulcal fundi [2], which are believed to be the geometric boundaries of sulcal banks. It has also been suggested that cortical thickness is a good marker of cytoarchitecture [3]. For instance, it has been reported that the anterior bank of the central sulcus, corresponding to the primary motor cortex, is thicker than the posterior bank of the central sulcus, corresponding to somatosensory area [4]. Despite previous separate studies in sulcal fundi extraction [5], cortical thickness measurement [4], and whole cortex parcellation [6], investigation of finer granularity parcellation of sulcal regions into sulcal banks and their correlations with the cytoarchitectonic signature of cortical thickness is still in its infancy. A systematic elucidation of the relationship between cortical geometry and cortical thickness can potentially contribute to better

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 108–117, 2010.

understanding of normal brain architecture and alternations in brain aging and brain diseases. Therefore, we were motivated to develop a robust and effective method to extract sulcal banks so that we can systematically study the cortical thickness on opposing sulcal banks. As the central sulcus (CS) is one of most reliable and important anatomic landmark on the cortical surface, we use CS as a test bed for algorithm development and evaluation using a large population of subjects. The proposed sulcal bank segmentation method consists of two major steps: parcellation of the cortical surface into sulcal basins and segmentation of parcellated sulcal basins into sulcal banks. For the first step, we employ the cortical sulcal parcellation method proposed in [7]. We concentrate on the second step: segmentation of the extracted sulcal basin into opposing sulcal banks, which is treated as a graph partition problem, which will be described in the following section.

2 Methods

Given a parcellated triangular sulcal basin, the sulcal bank segmentation method consists of two steps: 1) rough segmentation of the sulcal basin into two opposing sulcal banks using the normalized cuts (Ncuts) method [8]; 2) further refinement of the boundary using the graph cuts (Gcuts) method [9]. Fig. 1 provides an example of the sulcal bank segmentation method. Please note that the boundary between opposing sulcal banks can be considered as the sulcal fundus.

2.1 Normalized Cuts for Rough Sulcal Bank Segmentation

To adopt the Ncuts method [8] for sulcal bank segmentation, the triangular mesh of a sulcal basin is represented by an undirected weighted graph $\mathbf{G}_1 = (\mathbf{V}, \mathbf{E}_1)$, where $\mathbf{V}$ is the collection of the vertices in the sulcal basin, and $\mathbf{E}_1$ is the collection of the edges in the graph, which are formed between each pair of vertices in the sulcal basin. In the graph, the weight $w(i, j)$ on each edge represents the similarity between two vertices i and j. The Ncuts criterion for partitioning the graph into two disjoint parts $\mathbf{A}$ and $\mathbf{B}$ is defined as:

$$Ncut(\mathbf{A},\mathbf{B}) = \frac{cut(\mathbf{A},\mathbf{B})}{assoc(\mathbf{A},\mathbf{V})} + \frac{cut(\mathbf{A},\mathbf{B})}{assoc(\mathbf{B},\mathbf{V})} \tag{1}$$

where $cut(\mathbf{A},\mathbf{B}) = \sum_{u\in\mathbf{A},v\in\mathbf{B}} w(u,v)$, and $assoc(\mathbf{A},\mathbf{V}) = \sum_{u\in\mathbf{A},t\in\mathbf{V}} w(u,t)$. The $assoc(\mathbf{B},\mathbf{V})$ is defined similarly. The advantage of the Ncuts criterion is that it measures both the total dissimilarity between the different groups of data and the total similarity within the same groups without bias. In addition, the Ncuts method requires neither an explicit representation of the groups in data nor all vertices within a group being similar to a single representative vertex. These features are quite suitable for sulcal bank segmentation, since sulcal banks might be curved patches (Fig. 1) and an explicit representation of a sulcal bank by a parametric model or by a single representative vertex might not be appropriate.

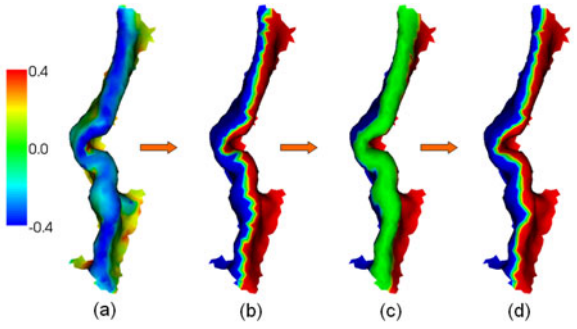


Fig. 1. An example of the proposed sulcal bank segmentation method. The sulcal basin is viewed from inside of the cortical surface for the convenience of inspection. (a) A central sulcal basin with the maximum principal curvature map. The color bar is on the left. (b) The sulcal bank segmentation results using Ncuts. The blue and red colors represent extracted two opposing sulcal banks and the green color represents the boundary between them. (c) The distinguished three parts. The green color represents the adjacent sulcal fundi region and the red and blue colors indicate two partial sulcal banks. (d) The final sulcal bank segmentation results.

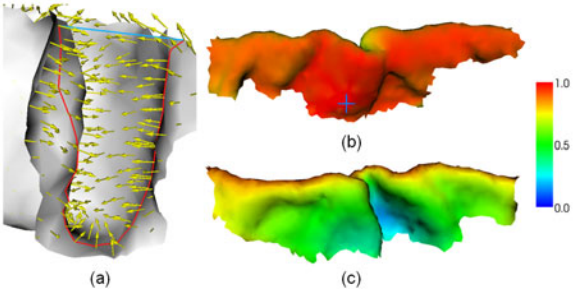


Fig. 2. An illustration of the similarity definition. (a) A cross-section of a central sulcal basin. The yellow arrows represent the normal directions. The red curve is the geodesic path between two vertices on opposing sulcal banks, and the blue line is the Euclidean distance. (b) and (c): the similarity map of a vertex labeled by the blue cross on a sulcal bank to all vertices on two sulcal opposing basins. The color bar is on the right.

Defining the appropriate similarity metric that captures the essence of the problem of sulcal bank segmentation is vital. According to our extensive observations, typically, a sulcal basin consists of two opposing sulcal banks that meet at the sulcal fundus, and the cross-section of a sulcal basin appears as a "V" or "U" shape. For instance, the central sulcal basin consists of anterior and posterior sulcal banks, which meet at the central sulcal fundus. In principal, the similarity between two vertices within the same sulcal bank should be large and the similarity between two vertices in opposing sulcal banks should be small. The similarity $w(i, j)$ is defined as a linear weighed combination of two similarity terms:

$$w(i, j) = \alpha \times w_d(i, j) + (1 - \alpha) \times w_n(i, j) \tag{2}$$

where α ($0 \le \alpha \le 1$) is a parameter used to control the trade-off between the distance similarity $w_d(i, j)$ and the angular similarity $w_n(i, j) \cdot w_n(i, j)$ is defined as:

$$w_n(i, j) = (1 + \mathbf{n}_i \cdot \mathbf{n}_j)/2 \tag{3}$$

where $\mathbf{n}$ is the normal direction. The distance similarity $w_d(i, j)$ is defined as the ratio between the Euclidean distance $d_e(i, j)$ and the geodesic distance $d_g(i, j)$:

$$w_d(i, j) = d_e(i, j)/d_g(i, j) \tag{4}$$

The range of $w(i, j)$ is: $0 \le w(i, j) \le 1$. The geodesic distance is the shortest path connecting two vertices along the triangular mesh of a sulcal basin. The basic idea is: if two vertices are within the same sulcal bank, the angular similarity of the two normal directions will be large and the ratio between the Euclidean distance and the geodesic distance is close to 1. Otherwise, the angular similarity will be low, and the ratio between the Euclidean distance and the geodesic distance is small. α is set to be 0.8 in this paper. Fig. 2 shows an illustration of the similarity definition for sulcal bank segmentation in Ncuts, which above criterion is approximated by solving a generalized eigenvalue problem [8].

2.2 Graph Cuts for Fine Sulcal Bank Segmentation

Since Ncuts is an approximate spectral technique for graph partition, the boundaries between the extracted sulcal banks could be inaccurate. Figure 1b shows an example. Therefore, we need to refine the sulcal bank segmentation result to obtain the accurate boundary between opposing sulcal banks. When refining the segmentation result, we impose hard constraints by automatically specifying certain vertices that have to be parts of a sulcal bank A or B. Meanwhile, we define a clear energy function for the segmentation. Thus, the rest of vertices are segmented as either parts of sulcal bank A or B by computing an optimum of the energy function that satisfies the hard constraints. To define the hard constraints, the sulcal bank segmentation result by Ncuts is divided into three parts: a part of sulcal bank A called A', a part of sulcal bank B called B', and the adjacent sulcal fundi region, which is defined as all of the two ring neighborhood vertices of the boundary between two segmented sulcal banks by Ncuts. The remaining two disjoint parts in the sulcal basin are considered as A' and B', which are treated as hard constraints, thus the feasible segmentation can only happen within the banded adjacent sulcal fundi region. Figure 1c shows an example of the distinguished three parts in a sulcal basin. Adding hard constraints helps obtain desirable segmentation results.

Let $L = (L_1, \ldots, L_p, \ldots L_N)$ be a binary vector and L_p represents the label assigned to vertex p in a sulcal basin. Each L_p can be either 0 or 1 (indicating sulcal bank A or B respectively). The energy function for sulcal bank segmentation is defined as:

$$E(L) = \lambda \cdot R(L) + B(L) = = \lambda \cdot \sum_{p \in P} R_p(L_p) + \sum_{p,q \in \varepsilon} B_{p,q}(L_p, L_q) \tag{5}$$

where $\mathcal{E}$ is the neighborhood system on the triangular sulcal basin. $R(L)$, called the regional term, measures the cost of assigning the label L_p to the vertex p. $R_p(L_p)$ should be small value when assigning vertex p to the sulcal bank which it belongs to. If a vertex is in the adjacent sulcal fundi region, the regional term is defined as:

$$R_p(L_p = l) = -\ln(\sum_i w(i,p) \cdot \delta(L_i = l) / \sum_i \delta(L_i = l)) \tag{6}$$

where $w(i,p)$ is the similarity between vertices i and p defined in Ncuts. If $L_i = l$, $\delta(L_i = l) = 1$; otherwise, $\delta(L_i = l) = 0$. $l \in \{0,1\}$ correspond to sulcal bank A and B respectively. $B(L)$, called the boundary term, measures the cost of assigning the label L_p, L_q to adjacent vertices p,q and is used to impose spatial smoothness. $B_{p,q}(L_p, L_q)$ should be large value when two vertices are within the same sulcal bank. To define the boundary term, we adopt the maximum principal curvature c, which are large negative values at sulcal bottoms and are large positive values at gyral crests [7]. The basic idea is: if the two vertices are at different sides of a sulcal fundus, where the two maximum principal curvatures will be large negative values since these two vertices are in sulcal bottoms, the boundary term should be a small value; otherwise, and the boundary term should be a large value. The boundary term $B_{p,q}(L_p, L_q)$ is defined as:

$$B_{p,q}(L_p, L_q) = (\exp(c(p)) + \exp(c(q))) / 2 \cdot \delta(L_p \neq L_q) \tag{7}$$

If $L_p \neq L_q$, $\delta(L_p \neq L_q) = 1$; otherwise, $\delta(L_p \neq L_q) = 0$. The parameter λ determines the relative contribution between the region term and the boundary term in the energy function. λ is set to be 30.0 in this paper. The optimal segmentation is the $\hat{L}$ that minimizes the energy function and meanwhile satisfies the imposed hard constraints.

Table 1. The definitions of graph edge weights

Edge	Weight	Condition
$\{p,q\}$	$B(p,q)$	$\{p,q\} \in \mathcal{E}$
$\{p,\mathbf{S}\}$	$\lambda \cdot R_p(B')$	$p \notin (A' \cup B')$
	$1 + \max_{p \in P} \sum_{\{p,q\} \in \mathcal{E}} B_{\{p,q\}}$	$p \in A'$
	0	$p \in B'$
$\{p,\mathbf{T}\}$	$\lambda \cdot R_p(A')$	$p \notin (A' \cup B')$
	0	$p \in A'$
	$1 + \max_{p \in P} \sum_{\{p,q\} \in \mathcal{E}} B_{\{p,q\}}$	$p \in B'$

To adopt Gcuts method for sulcal bank segmentation, a graph is defined as $\mathbf{G}_2 = (\mathbf{V}_2, \mathbf{E}_2)$, where $\mathbf{V}_2$ is the collection of the nodes, and $\mathbf{E}_2$ is the set of edges. $\mathbf{V}_2 = \mathbf{V} \cup \{\mathbf{S}, \mathbf{T}\}$, where $\mathbf{V}$ is set of vertices in the sulcal basin, $\mathbf{S}$ and $\mathbf{T}$ represent two terminal nodes: source and sink, which stand for A' and B' respectively. $\mathbf{E}_2 = \mathbf{E}_n \cup \mathbf{E}_t$, where $\mathbf{E}_n$ and $\mathbf{E}_t$ represent two collections of edges: n-links and t-links. The edges between adjacent vertices are called n-links. And the edges connecting vertices to two terminals are called t-links. Each vertex p has two t-links $\{\mathrm{p}, \mathbf{S}\}$ and $\{\mathrm{p}, \mathbf{T}\}$ connecting it to two terminals. Each pair of neighboring vertices $\{\mathrm{p}, \mathrm{q}\}$ is connected by an n-link. The weight of edges in the graph $\mathbf{G}_2$ is defined in the Table 1. It has been proved that the segmentation defined by the minimum cut of the above constructed graph minimizes Eq. (5) among all segmentations satisfying the hard constraints [9]. In our implementation, we adopt the max-flow graph cut algorithm in [10]. Figure 1 shows an example of the sulcal bank segmentation.

3 Results

The publicly available OASIS neuroimaging dataset [11] is used to evaluate the proposed method. The sulcal bank segmentation is performed on the gray/white matter surfaces. All of the topologically correct and geometrically accurate cortical surfaces used here are reconstructed via the method in [12]. We have successfully reconstructed 414 cortical surfaces from the OASIS dataset. After parcellation of cortical surfaces into sulcal basins using the method in [7], we obtained 401 correctly extracted central sulcal basins on both hemispheres. We found that the sulcal basin segmentations are quite reasonable. The central sulcal basins are interactively identified from the parcellated cortical surfaces by experts. All of the correctly extracted central sulcal basins are segmented into anterior and posterior sulcal banks using the presented method in this paper. We extract the anterior and posterior sulcal banks of 401 left and 401 right central sulci from the OASIS dataset [11]. Our inspection of 60 randomly selected central sulci shows visually correct results. Figure 3 shows the sulcal bank segmentation of 15 randomly selected central sulci.

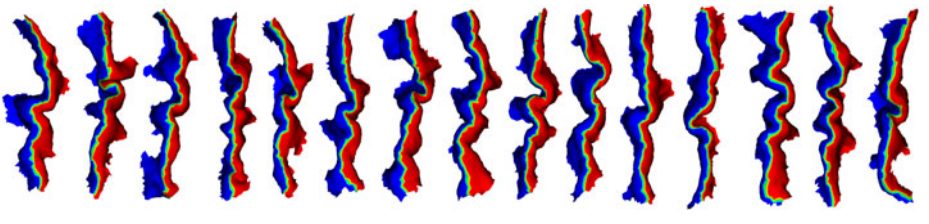


Fig. 3. The sulcal bank segmentation results on 15 randomly selected central sulci

To quantitatively evaluate the sulcal bank segmentation results, we have two experts who manually annotated the anterior and posterior central sulcal banks of both hemispheres of 15 randomly selected subjects. We applied the Dice coefficient between automatically extracted and manually labeled sulcal banks to validate the

proposed method. The Dice coefficients of the anterior and posterior sulcal banks on both hemispheres for the 15 subjects are above 0.92 for both experts, indicating the relatively accurate performance of our method. Figure 4 shows the details of the Dice coefficients on both hemispheres of the 15 subjects, in comparison with expert 1.

4 Applications

With the sulcal bank segmentation results, we are able to quantitatively study the cortical thickness on anterior and posterior sulcal banks of central sulci. We adopt the method in [4] to calculate the cortical thickness on all of the extracted central sulcal banks in the OASIS dataset [11]. Over all, we have 401 central anterior and posterior sulcal banks on each hemisphere of 307 healthy subjects. In the 307 subjects, on right hemispheres, the average cortical thicknesses of the central anterior and posterior sulcal banks are 2.66±0.28mm and 2.30±0.30mm, respectively. On left hemispheres, the average cortical thickness of the central anterior and posterior sulcal banks is 2.81±0.34mm and 2.30±0.29mm, respectively. Figure 5 and 6 show the detailed distributions of the average cortical thickness of the central anterior and posterior sulcal banks of the 307 subjects on right and left hemispheres, respectively. These figures also show the differences between average cortical thickness of the central anterior sulcal bank and that of the corresponding central posterior sulcal bank on both hemispheres. The t-test is applied to these results. Considering that $P<0.001$ on both hemispheres, we conclude that the anterior sulcal bank is significantly thicker than the posterior sulcal bank of central sulci on both hemispheres. These results are consistent with previous reports on imaging studies [4], but with much larger number of subjects and in a more systematic fashion.

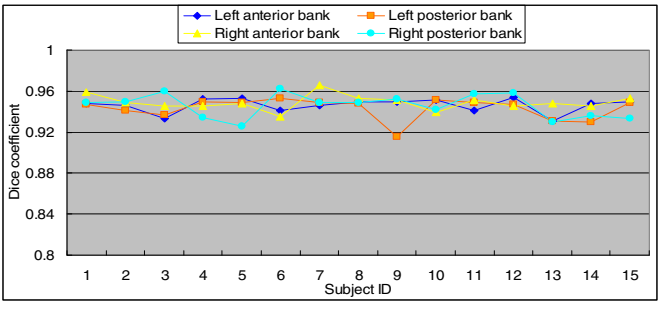


Fig. 4. The Dice coefficients compared to expert 1 for the anterior and posterior central sulcal banks of the 15 subjects

It has been reported that the cortex was generally thicker in the left hemisphere than the right hemisphere, and significant leftward asymmetry has been identified in several cortical regions [13]. We are interested in inspecting whether there exit a significant leftward asymmetry in these two opposing sulcal banks of central sulci. In the 307 normal subjects, we find 301 subjects with successfully extracted both left and right central sulcal banks. The average cortical thickness of left and right central

anterior sulcal banks are 2.81±0.34mm and 2.66±0.28mm, respectively. Figure 7 shows the detailed distributions of the average cortical thickness of the left and right central anterior sulcal banks of the 301 subjects. The figure also shows the differences between average cortical thickness of the left central anterior sulcal bank and that of the corresponding right central anterior sulcal bank. The t-test is applied to these results. Considering that $P<0.001$, we conclude that the central anterior sulcal bank is significantly thicker at left side. The average cortical thickness of left and right central posterior sulcal banks are 2.30±0.29mm and 2.30±0.30mm, respectively. No significant asymmetry has been found on central posterior sulcal banks.

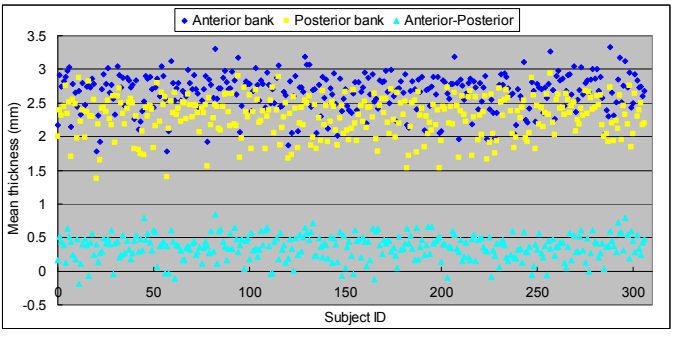


Fig. 5. The average GM thickness distributions of the anterior and posterior sulcal banks of the central sulci on right hemispheres of 307 subjects

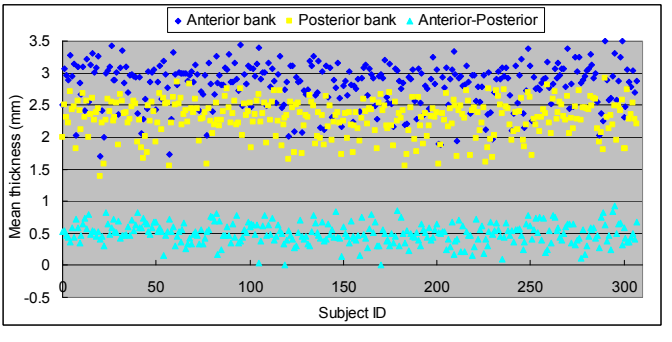


Fig. 6. The average GM thickness distributions of the anterior and posterior sulcal banks of the central sulci on left hemispheres of 307 subjects

Many cross-sectional studies have reported thinner cortices in older than younger individuals [14, 15], especially for the central sulcus. Therefore, we studied the age-related trend of the average cortical thickness of the anterior and posterior sulcal banks of the central sulcus using second order polynomial regression analysis. The correlation of the anterior and the posterior banks of the right central sulcus are $R=-0.7127$ and $R=-0.7851$, respectively. And the correlation of the anterior and the posterior banks of the left central sulcus are $R=-0.7915$ and $R=-0.7900$, respectively.

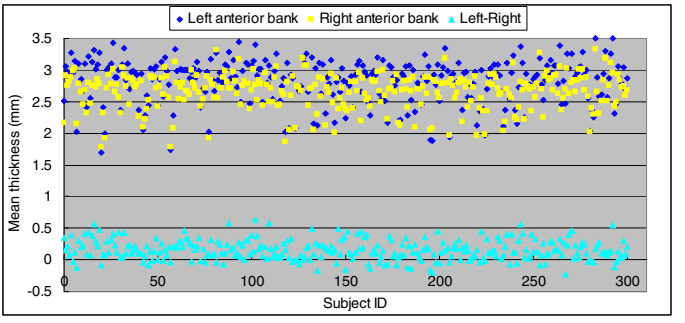


Fig. 7. The distributions of average GM thickness of the left anterior and right anterior sulcal banks of the central sulci of 301 subjects without dementia

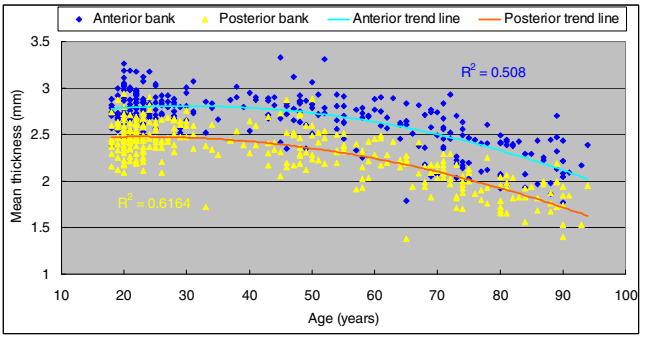


Fig. 8. The distributions and trend lines of average cortical thickness of right anterior (blue) and posterior sulcal bank (yellow) of the central sulcus along with the ages

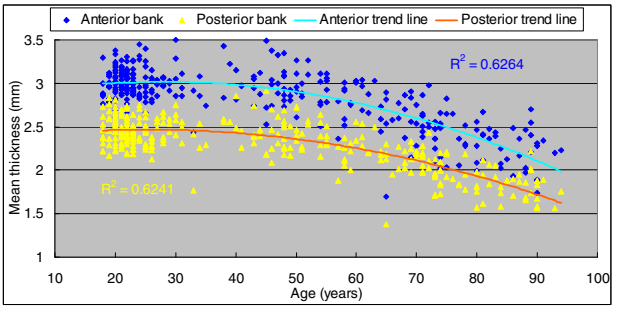


Fig. 9. The distributions and trend lines of average cortical thickness of left anterior (blue) and posterior sulcal bank (yellow) of the central sulcus along with the ages

Figure 8 and 9 show the distributions and trend lines of mean cortical thickness of anterior and posterior sulcal banks of central sulci along with the ages of 307 healthy subjects on right and left hemispheres, respectively. All these results indicate a gradual decrease in cortical thickness across the adult life span of both the anterior

and the posterior banks of both the left and right central sulci of 307 healthy subjects. These results support the report that thinning of cortical thickness was found in the central sulcus [14, 15].

5 Conclusion

In this paper, we presented a novel method of geometric similarity based graph partition for sulcal bank segmentation. As a test bed application, we applied the segmentation method to study the cortical thickness of anterior and posterior sulcal banks of the central sulcus in the OASIS dataset and obtained reasonable results. Our results support that cortical geometric characteristic is a good predictor of cortical cytoarchitecture. In the future, we plan to apply the method to other major sulci in normal brains, in order to systematically elucidate the relationships between sulcal banks and the cortical thickness.

References

1. Fischl, B., Rajendran, N., Busa, E., et al.: Cortical folding patterns and predicting cytoarchitecture. Cereb. Cortex 18(8), 1973–1980 (2008)
2. Welker, W.: Why does cerebral cortex fissure and fold? A review of determinants of gyri and sulci. Cereb. Cortex 8b (1990)
3. Lerch, J.P., Worsley, K., Shaw, W.P., et al.: Mapping anatomical correlations across cerebral cortex (MACACC) using cortical thickness from MRI. NeuroImage 31(3), 993–1003 (2006)
4. Fischl, B., Dale, A.M.: Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc. Natl. Acad. Sci. 97(20), 11050–11055 (2000)
5. Li, G., Guo, L., Nie, J., Liu, T.: An automated pipeline for cortical sulcal fundi extraction. Medical Image Analysis 14(3), 343–359 (2010)
6. Fischl, B., van der Kouwe, A., Destrieux, C., et al.: Automatically parcellating the human cerebral cortex. Cereb. Cortex 14(1), 11–22 (2004)
7. Li, G., Guo, L., Nie, J., Liu, T.: Automatic cortical sulcal parcellation based on surface principal direction flow field tracking. NeuroImage 46(4), 923–937 (2009)
8. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 22(8), 888–905 (2000)
9. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient N-D image segmentation. International Journal of Computer Vision 70(2), 109–131 (2006)
10. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell. 23(11), 1222–1239 (2001)
11. Marcus, D.S., Wang, T.H., Parker, J., et al.: Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J. Cogn. Neurosci. 19(9), 1498–1507 (2007)
12. Liu, T., Nie, J., Tarokh, A., et al.: Reconstruction of central cortical surface from brain MRI images: method and application. NeruoImage 40(3), 991–1002 (2007)
13. Luders, E., Narr, K.L., Thompson, P.M., et al.: Hemispheric asymmetries in cortical thickness. Cereb. Cortex 16(8), 1232–1238 (2006)
14. Salat, D.H., Buckner, R.L., Snyder, A.Z., et al.: Thinning of the cerebral cortex in aging. Cereb. Cortex 14(7), 721–730 (2004)
15. Rettmann, M.E., Kraut, M.A., Prince, J.L., Resnick, S.M.: Cross-sectional and longitudinal analyses of anatomical sulcal changes associated with aging. Cereb. Cortex 16(11), 1584–1594 (2006)

A Framework for 3D Analysis of Facial Morphology in Fetal Alcohol Syndrome

Jing Wan[1,2,4], Li Shen[1,2,⋆], Shiaofen Fang[1], Jason McLaughlin[1], Ilona Autti-Rämö[5], Åse Fagerlund[6], Edward Riley[7], H. Eugene Hoyme[8], Elizabeth S. Moore[9], Tatiana Foroud[3,⋆], and Collaborative Initiative on Fetal Alcohol Spectrum Disorders (CIFASD)

[1] Dept. of Computer and Info. Science, Purdue University, Indianapolis, IN, USA
shenli@iupui.edu
[2] Dept. of Radiology and Imaging Sciences,
[3] Dept. of Medical and Molecular Genetics,
Indiana University School of Medicine, Indianapolis, IN, USA
tforoud@iupui.edu
[4] Jiangxi University of Finance and Economics, Jiangxi, China
[5] Dept. of Child Neurology, HUCH Hospital for Children & Adolescents, Findland
[6] Folkhälsan Research Center, Helsinki, Finland
[7] Department of Psychology, San Diego State University, CA, USA
[8] Sanford School of Medicine, University of South Dakota, Vermillion, SD, USA
[9] St. Vincent Women's Hospital, Indianapolis, IN, USA

Abstract. Surface-based morphometry (SBM) is widely used in biomedical imaging and other domains to localize shape changes related to different conditions. This paper presents a computational framework that integrates a set of effective surface registration and analysis methods to form a unified SBM processing pipeline. Surface registration includes two parts: surface alignment in the object space by employing the iterative closest point (ICP) method, and surface alignment in the parameter space by using conformal mapping and landmark-based thin-plate spline methods. Statistical group analysis of registered surface data is then conducted by surface-based general linear model and random field theory addressing multiple testing issues. The effectiveness of the proposed framework is demonstrated by applying it to a fetal alcohol syndrome (FAS) study for identifying facial dysmorphology in FAS patients.

1 Introduction

The adverse effects of alcohol on the developing fetus fall along a continuum. The collection of these disorders is known as fetal alcohol spectrum disorders (FASD). Fetal alcohol syndrome (FAS), considered as a more severe subset of FASD, can be defined by recognizable facial dysmorphology, growth deficits, and behavioral problems [6]. Individuals diagnosed with FAS present a pattern

⋆ Correspondence to L. Shen or T. Foroud. Supported in part by NIAAA U01 AA014809 and NIBIB R03 EB008674.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 118–127, 2010.

of minor facial anomalies including short palpebral fissures, flat nasal bridge, smooth philtrum, thin upper lip, and flat mid-face [12]. The ongoing clinical challenge is to expand the recognizable facial features so more individuals with prenatal alcohol exposure that do not exhibit the classic facial phenotype can be identified, allowing for early interventions.

Craniofacial anthropometry has been used to accurately identify individuals with FAS [13]. However, anthropometric assessments can be time consuming and usually require an experienced anthropometrist to obtain the measurements. So there is a need for newer techniques, which in combination with a clinician's assessment, would provide rapid and accurate pre-screening and early diagnosis of children with a FASD [6]. Three-dimensional (3D) surface-based morphometry offers the opportunity to solve the problem.

Surface-based morphometry (SBM), widely used in biomedical imaging to study various structures of interest, is used to identify morphometric abnormalities associated with a particular condition, assisting with diagnosis and treatment. Many studies [3,9] of facial morphology focus on the delineation of characteristic features or building computational models of face-shape variation. Hammond et al. [9] proposed a dense surface model of the huaman face. They used 3D thin-plate spline (TPS) to align the landmarks of each face, employed iterative closest point (ICP) method to build correspondence by taking the closest point on each surface from each vertex on a base mesh, and then applied the inverse of the TPS warp to map each surface back to its original location. The base mesh was chosen as an individual from the data and thus not a regular mesh. Using such a base mesh to sample other meshes may not be ideal for SBM which typically works on a uniformly sampled surface manifold. In addition, since 3D TPS has no analytical inverse, the error introduced by approximating the TPS inverse and its computational complexity are unclear. Wang et al. [21] proposed a non-rigid 3D motion tracking algorithm using harmonic maps with added feature correspondence constraints to build dense one-to-one inter-frame correspondences. This method may not be applicable to data sets where only geometric information is available or surface correspondence cannot be implicated by texture information.

Here we propose a novel computational framework that performs SBM on 3D facial imaging data and demonstrate its effectiveness via an FAS application. We are given two groups of 3D facial surfaces represented by triangular meshes. Each mesh is assumed to have a disk topology and does not contain any hole. A set of landmarks are available on each mesh to pre-define a coarse correspondence between surfaces. Our goal is to localize morphological changes related to the group condition (i.e., FAS versus controls). Our SBM framework includes three key components: (1) Surface alignment in the object space is achieved by employing the ICP algorithm. (2) Surface registration in the parameter space is obtained by using conformal mapping and landmark-based 2D TPS methods. (3) Statistical group analysis of registered surface data is conducted directly on the surface manifold to avoid distortion introduced by surface flattening, and a surface-based general linear model with random field theory is used to achieve this goal and address multiple testing issues.

Our approach sacrifices full automation by using landmarks in exchange for more robustness and higher accuracy to surface registration, allowing users to define critical locations where surface correspondence should be established. Instead of directly working in 3D domain like [9], our approach establishes the surface correspondence via a 2D parametric domain, making the entire processing pipeline simple, efficient, and easy to implement and interpret. While this pipeline is demonstrated in an FAS study, it is a general framework and could be applied to other 3D surface objects with disk-like topology and landmarks.

2 Materials and Methods

2.1 Data Set

The data set used in this study included 44 (18 males and 26 females, age 13.0 ± 3.6) FAS and 52 (20 males and 32 females, age 13.9 ± 3.6) healthy control (HCL) participants from Helsinki, Finland. The participants or their legal guardians provided written informed consents. Each participant was examined and classified as FAS, no FAS, or deferred according to a standardized assessment [11]. Preliminary diagnosis was determined on the basis of facial structural features and growth deficiency consistent with the revised Institute of Medicine criteria [10]. The final diagnoses were considered with alcohol exposure data which were collected through a standard questionnaire consisting of four questions in the interview or from a review of available study data. This study only included the individuals determined as FAS with prenatal alcohol exposure and the ones designated as no FAS without prenatal alcohol exposure.

A standard scheme was utilized to collect 3D facial images using Minolta Vivid 910 laser scanners. Each participant was seated approximately 660 mm from the scanner and six scans were taken as: two frontal, two 45 degree to the right of the frontal axis, and two 45 degree to the left of the frontal axis. These three directions of scans ensured the entire facial area was covered. A stitching processing using a commercially available software package, Rapidform 2004, was applied to merge the scans of the three views into one single 3D surface image [6]. 33 landmarks (Figure 1) were then manually defined by a trained technician. The landmarks were prominent and easily identifiable points on each face (e.g. the corners of the eyes). They provide meaningful constraints for regions of interest in facial dysmorphology to guide the registration procedure.

2.2 Surface Registration

We first register all the surfaces in the object space. The iterative closest point (ICP) algorithm [1] is used to register each surface to a template surface, which is pre-selected as an HCL surface. This rigid transformation normalizes the orientation and location of each surface. After that, surface correspondence is established via 2D parameterization, where conformal mapping and 2D thin-plate spline (TPS) [2] are employed. Our goal is to achieve a smoothing mapping

between any surface and the 2D parameter domain, and assign the same landmark of different models with the same location in the parameter space.

We first describe how conform mapping is implemented. To perform statistical shape analysis on 3D facial surfaces, we need to establish a meaningful correspondence between them. One way to achieve the goal is to map these facial surfaces into a standard space while preserving geometric information on the original structures as much as possible [7,8]. So this becomes a surface parameterization problem.

Surface parameterization, defined as one-to-one mapping from a surface into another parameter domain, can always introduce distortion in either angles or areas [7]. A good mapping is the one which minimizes the distortions to some extent. To achieve this goal, conformal mapping is one way to minimize the angular distortion [8]. Riemann theorem states that conformal mapping of a smooth surface into a plane exists for any simply-connected plane domain [4]. Since meshes of a smooth surface can be viewed as approximations of the surface, it is possible to map them to a plane with very little angular distortion [15].

Before describing the comformal mapping algorithm used in our framework, we first introduce some basic concepts. Suppose a surface $M_1 \subset \mathbb{R}^3$ has the parametric representation $\mathbf{x}(u^1, u^2) = (x_1(u^1, u^2), x_2(u^1, u^2), x_3(u^1, u^2))$ for points (u^1, u^2) in some domain in $\mathbb{R}^2$. The first fundamental form of M_1 is

$$ds_1^2 = \sum_{ij} g_{ij} du^i du^j \quad \text{where} \quad g_{ij} = \frac{\partial \mathbf{x}}{\partial u^i} \cdot \frac{\partial \mathbf{x}}{\partial u^j} \ (i, j = 1, 2) \tag{1}$$

Another plane $M_2 \subset \mathbb{R}^2$ is similarly represented by $\widetilde{\mathbf{x}}(\tilde{u}^1, \tilde{u}^2)$. Define a mapping $f : M_1 \mapsto M_2$ between two surfaces. If f is a conformal mapping, then there is some scalar function $\eta \neq 0$, such that $ds_1^2 = \eta(\tilde{u}^1, \tilde{u}^2)((d\tilde{u}^1)^2 + (d\tilde{u}^2)^2)$. As shown in [7], two Laplace's equations are obtained as

$$\triangle_s \tilde{u}^1 = 0, \ \triangle_s \tilde{u}^2 = 0 \tag{2}$$

$\triangle_s$ is the Laplace-Beltrami operator, which can be written as $\triangle_s = \mathrm{div}_s \mathrm{grad}_s$. To find the solution to Equation (2), the conformal mapping f can be viewed as minimizing the Dirichlet energy:

$$E_0(f) = \frac{1}{2} \int_s \|\mathrm{grad}_s f\|^2 \tag{3}$$

To compute f, Eck et al. [5] proposed an approach, called discrete harmonic map, which extends the graph embedding method of Tutte [19]. In their method, the boundary vertices of the meshes are first mapped to the boundary of the unit disk. Then the positions of the remaining vertices can be computed by solving equations:

$$L\tilde{u}^1 = 0, \qquad L\tilde{u}^2 = 0 \tag{4}$$

$$L_{ij} = \begin{cases} -\sum_{k \neq i} L_{ik}, & i = j \\ w_{ij}, & (i, j) \in E \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

$$w_{ij} = \cot \alpha_{ij} + \cot \beta_{ij}, \tag{6}$$

where α_{ij} and β_{ij} are the opposite angles in the two triangles sharing an edge (i, j) [15].

This study employs Eck's method to perform conformal mapping from 3D facial meshes to 2D meshes in the unit disk. A public matlab toolbox, Toolbox Graph [14], is used to implement the algorithm. Here we perform conformal mapping to map each vertex of a 3D mesh to the corresponding 2D position in a unit disk with fixed boundary. Suppose an individual is represented as a set of vertices (Xa, Ya, Za). After applying conformal mapping Φ, which is bijective, each individual gets new coordinates $(\Phi(Xa), \Phi(Ya))$ in the 2D disk domain. As an example, let the height (i.e., Z value of each vertex in the original model) be the geometric information of our interest, and thus the mapped individual can be represented as $(\Phi(Xa), \Phi(Ya), Za)$. Of note, other surface features (e.g., x, y coordinates, curvature, etc.) can be studied in the same way. For simplicity, we explain our method here by focusing on z coordinates only.

After applying conformal mapping to all the subjects in the data set, we pick a healthy individual as the template and then register all the subjects to the template using landmark based thin-plate splines (TPS) warp Ψ [2] in the 2D domain. Now the landmarks of each individual are exactly aligned to those of the template. The remaining parts of the individuals are interpolated according to the movement of their landmarks. An individual can then be represented as $(\Psi(\Phi(Xa)), \Psi(\Phi(Ya)), Za)$, where Ψ denotes the TPS registration function.

Since each individual mesh has different number of vertices and triangles, we need to resample them using a regular mesh grid defined in the disk. After resampling, all the individuals in the data set have the same mesh topology and can be compared with each other. Z-coordinates (i.e., heights) of the re-sampled points can be obtained by using cubic interpolation for each individual, and are used to extract surface signals in subsequent analyses.

2.3 Statistical Shape Analysis

After surface registration, all facial surfaces are aligned to the same reference system. This facilitates the subsequent analysis on the facial surfaces, including extracting surface signals and performing statistical inference on the surface manifold using general linear model (GLM) [17,20]. We use X_t to denote the atlas, which is computed as an average of all registered healthy controls (HCLs). For an individual surface X, we use its deformation field $\delta(X) = X - X_t$ as surface signals to describe the shape based on the atlas X_t. In this study, we examine two types of surface shape signals: (1) deformation scalar along z-axis direction and (2) mean curvature of each vertex.

Remember our goal is to detect significant shape changes on facial surfaces between HCL group and FAS group. We consider the following GLM

$$y = X\Psi + Z\Phi + \epsilon \tag{7}$$

where the dependent variable y is our surface signal; $X = (x_1, \cdots, x_p)$ are the variables of interest such as Group; $Z = (z_1, \cdots, z_k)$ are the variables whose

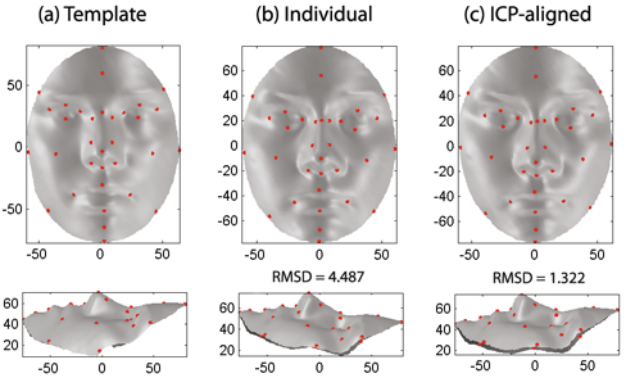


Fig. 1. Registration in the object space by aligning an individual to the template using ICP: (1) template, (2) individual, and (3) ICP-aligned individual

effects we want to exclude, such as Age and Gender; and $\Psi = (\psi_1, \cdots, \psi_p)^T$ and $\Phi = (\phi_1, \cdots, \phi_k)^T$ are the coefficients; and ϵ is the error term. The goal is to test if X is significant (i.e. $\Psi \neq 0$) for $y \in \partial\Omega$, where $\partial\Omega$ indicates the surface manifold. We use SurfStat to test our GLMs. SurfStat is a Matlab toolbox for statistical analysis of univariate and multivariate surface and volumetric data using linear mixed effects models and random field theory (RFT) [22]. Using SurfStat notation (coefficients excluded in the equation), we examine the following two models, where dependent variables (i.e. y in Eq.(7)) are our surface signals defined above, and RFT is used for multiple comparison correction.

Model	Description
M1 = 1 + Status	Diagnosis effect on surface signals
M2 = 1 + Status + Age + Gender	Diagnosis effect on surface signals with controlling Age and Gender effect

3 Experimental Results

3.1 Surface Registration

Figure 1 shows an example of registration result in the object space by aligning an individual to the template using ICP, which normalizes the orientation and location of the initial configuration. The root mean square distance (RMSD) between the individual and the template before ICP registration is 4.487 and it is reduced to 1.322 after ICP registration.

Figure 2 shows a sample registration procedure in the parameter space by aligning an individual's conformal map (d) to the template conformal map (c) using TPS. While landmarks are aligned perfectly between the individual and the template, a smooth mapping from the individual to the template is obtained for establishing the surface correspondence. To quantify the registration quality,

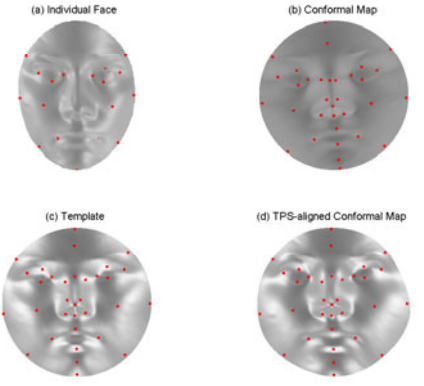


Fig. 2. Registration in the parameter space by aligning an individual's conformal map to the template conformal map using TPS: (1) individual face, (2) individual conformal map, (3) template conformal map, and (4) TPS-aligned individual conformal map

we consider two factors: (1) the area distort cost (ADC) (defined in [16]) from the object surface to the parameter domain (i.e., 2D disk), and (2) RMSD between landmarks of the individual and the template in the parameter domain. Our goal is to achieve RMSD=0 while controlling the area distortion ADC. If we just use conformal mapping, we have ADC of 1.1508 ± 0.0231 (mean±std) and RMSD of 0.6841 ± 0.1124 for all the subjects in our data. If we combine conformal mapping with TPS, we have ADC of 1.3723 ± 0.0987 and RMSD of 0 ± 0. In this case, although our ADC gets slightly increased, RMSD=0 guarantees that all the landmarks are perfectly aligned across all the subjects.

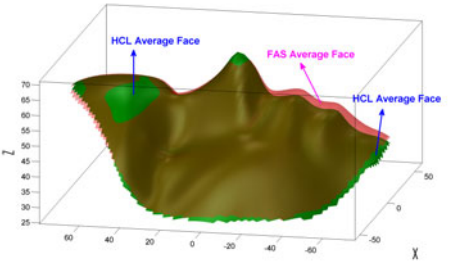


Fig. 3. HCL average face and FAS average face: HCL average face is used as the atlas for extracting surface signals

Figure 3 shows the average of all the HCL faces and the average of all the FAS faces. The HCL average face is used as the atlas for extracting surface signals. Figure 4 shows the surface registration and signal extraction results for one HCL and two FAS examples. The three facial surfaces after being aligned in object space are shown in Figure 4(a) and their registered surfaces after conformal

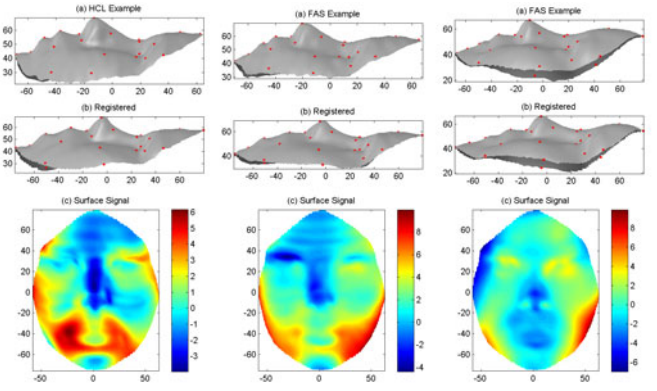


Fig. 4. One HCL and two FAS examples of surface registration and signal extraction results: (a) original surface, (b) registered surface, and (c) surface signals. Note that the x, y coordinates of landmarks after registration are consistent across subjects (see (b)). For each participant, the landmark z values (i.e., heights) stay the same before and after registration (see (a-b)).

mapping and TPS are shown in (b). Note that the x, y coordinates of landmarks after registration in parameter domain are consistent across subjects (see (b)). For each participant, the landmark z values (i.e., heights) stay the same before and after registration (see (a-b)). (c) shows the surface signals we extract from the three surfaces. The surface signals visualized here are deformation scalar along z-axis direction (i.e. height). Red color represents positive surface signal value while blue color denotes negative surface signal value.

3.2 Statistical Shape Analysis

As mentioned in Section 2.3, we examine diagnosis effect on surface signals between HCL and FAS, with and without involving age and gender as covariates in Model M_2 and Model M_1, respectively. Since the results of M_1 and M_2 are very similar, only M_2 results are shown in Figure 5: (a,c) the maps of the t statistics, (b,d) the maps of corrected P values for peak and clusters (only regions with $p \leq 0.01$ are color-mapped). Let us first look at the height results (a,b). As described earlier, the signal at each surface location is defined as the deformation from the atlas to an individual along z-axis direction. So the signals take either positive values for outward deformations or negative values for inward deformations. As the contrast of M_2 is defined as "FAS−HCL", the colorful T maps in (a) can be explained as follows: the shape of an FAS face tends to be contractive in the red regions and expansive in the blue regions compared to that of an HCL face. All the significant regions in (b) correspond to the red regions in (a), indicating FAS faces are expansive in those regions compared to HCL faces. When P threshold is set to 0.01, the significant regions cover the corner of two eyes, lips, and philtrum. These results are in accordance with previously known facial

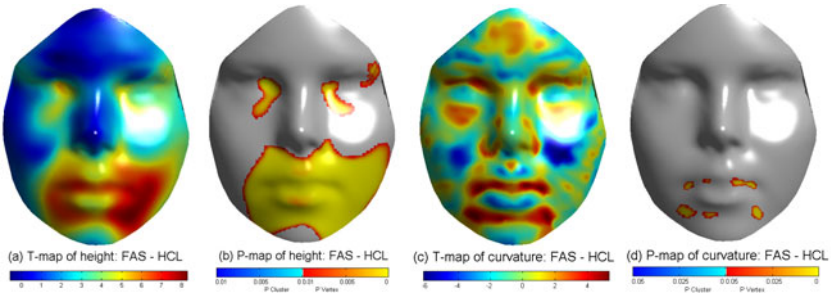


Fig. 5. Diagnosis effect (i.e., FAS-HCL) on surface signals (heights in (a,b), curvatures in (c,d)) while controlling for Age and Gender. (a,c) show the maps of the t statistics. (b,c) show the maps of corrected P values for peak and clusters with $p \leq 0.01$.

features for FAS diagnosis (e.g., short palpebral fissures, upper lip and smooth philtrum [10,18]). As to the curvature results (c,d), while few significant areas are identified in P-map (d), the T-map (c) shows a very interesting red region in the philtrum. Since curvatures in a philtrum region usually take negative values indicating concave, red color of that region in (c) denotes larger curvature values in FAS than in HCL and consequently smoother philtrum in FAS. This matches the existing findings.

4 Conclusions

In this paper, we presented a novel computational framework for surface based morphometry (SBM). It integrated a set of effective surface registration and analysis methods to form a unified SBM processing pipeline. Three parts were included in this pipeline: (1) surface alignment in the object space by performing the iterative closest point algorithm, (2) surface registration in the parameter space by employing conformal mapping and landmark-based thin-plate spline methods, and (3) statistical group analysis on registered facial surfaces by using general linear model with random field theory. This framework was applied to 3D analysis of facial morphology in fetal alcohol syndrome (FAS). The goal was to identify regional facial changes of the FAS group compared to healthy controls using 3D facial imaging data. The results identified facial dysmorphology patterns in FAS that were consistent with prior findings. This demonstrated the effectiveness of our framework. The framework is relatively simple and efficient. It can be applied to other applications dealing with similar data and derive results that are easy to interpret. One interesting future direction could be to examine additional surface signals such as tensors and displacement vectors. Another direction would be to apply this technique to individuals with known alcohol exposure that do not exhibit the classic FAS face.

References

1. Besl, P.J., McKay, N.D.: A method for registration of 3-d shapes. IEEE Trans. on Pattern Analysis and Machine Intelligence 14(2), 239–256 (1992)
2. Bookstein, F.L.: Shape and the information in medical images: a decade of the morphometric synthesis. Comput. Vis. Image Underst. 66(2), 97–118 (1997)
3. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Calculus of nonrigid surfaces for geometry and texture manipulation. IEEE Trans. on Visualization and Computer Graphics 13(5), 902–913 (2007)
4. do Carmo, M.P.: Differential geometry of curves and surfaces. Prentice Hall, Englewood Cliffs (1976)
5. Eck, M., DeRose, T.D., Duchamp, T., et al.: Multiresolution analysis of arbitrary meshes. In: Proc. of ACM SIGGRAPH 1995, pp. 173–182 (1995)
6. Fang, S., McLaughlin, J., et al.: Automated diagnosis of fetal alcohol syndrome using 3D facial image analysis. In: Orthodontics & Craniofacial Res., pp. 162–171 (2008)
7. Floater, M.S., Hormann, K.: Surface parameterization: a tutorial and survey. In: Advances in Multiresolution for Geometric Modelling, pp. 157–186. Springer, Heidelberg (2005)
8. Gu, X., Wang, Y., et al.: Genus zero surface conformal mapping and its application to brain surface mapping. IEEE Trans. on Medical Imaging 23(8), 949–958 (2004)
9. Hammond, P., Hutton, T.J., Allanson, J.E., et al.: 3D analysis of facial morphology. American Journal of Medical Genetics 126A, 339–348 (2004)
10. Hoyme, H., May, P., Kalberg, W., et al.: A practical clinical approach to diagnosis of fetal alcohol spectrum disorders: Clarification of the 1996 institute of medicine criteria. Pediatrics, 39–47 (2005)
11. Jones, K.L., Robinson, L.K., Bakhireva, L.N., et al.: Accuracy of the diagnosis of physical features of fetal alcohol syndrome by pediatricians after specialized training. Pediatrics 118, 1734–1738 (2006)
12. Jones, K.L., Smith, D.W.: Recognition of the fetal alcohol syndrome in early infancy. Lancet 2, 999–1001 (1973)
13. Moore, E.S., Ward, R.E., Wetherill, L.F., et al.: Unique facial features distinguish fetal alcohol syndrome patients and controls in diverse ethnic populations. Alcoholism: Clinical and Experimental Research 31, 1707–1713 (2007)
14. Peyre, G.: Toobox_Graph, http://www.ceremade.dauphine.fr/~peyre/matlab/graph/content.html
15. Sheffer, A., Praun, E., Rose, K.: Mesh parameterization methods and their applications. Found. Trends. Comput. Graph. Vis. 2(2), 105–171 (2006)
16. Shen, L., Farid, H., McPeek, M.A.: Modeling 3-dimensional morphological structures using spherical harmonics. Evolution 63(4), 1003–1016 (2009)
17. Shen, L., Saykin, A., Chung, M.K., et al.: Morphometric analysis of hippocampal shape in mild cognitive impairment: an imaging genetics study. In: IEEE 7th Int. Sym. on Bioinformatics and BioEngineering, Boston, MA, pp. 211–217 (2007)
18. Stoler, J.M., Holmes, L.B.: Recognition of facial features of fetal alcohol syndrome in the newborn. American Journal of Medical Genetics Part C 127(1), 21–27 (2004)
19. Tutte, W.T.: How to draw a graph. London Mathematical Soc. 13, 743–768 (1963)
20. Wan, J., Shen, L., Sheehan, K.E., et al.: Shape analysis of thalamic atrophy in multiple sclerosis. In: MIAMS, pp. 93–104. Imperial College, London (2009)
21. Wang, Y., Gupta, M., et al.: High resolution tracking of non-rigid 3d motion of densely sampled data using harmonic maps. In: Int. Conf. on Computer Vision (2005)
22. Worsley, K.J., Andermann, M., Koulis, M., et al.: Detecting changes in nonisotropic images. Human Brain Mapping 8, 98–101 (1999)

Feature Driven Rule Based Framework for Automatic Modeling of Organic Shapes in the Design of Personalized Medical Prosthetics

Sajjad Baloch[1], Konrad Sickel[2], Vojtech Bubnik[3], Rupen Melkisetoglu[1], Sergei Azernikov[1], Andreas Reh[4], Artem Boltyenkov[4], and Tong Fang[1]

[1] Siemens Corporate Research, Princeton, NJ, USA
[2] Friedrich-Alexander University, Erlangen, Germany
[3] Siemens Hearing Aid Instruments, Piscataway, NJ, USA
[4] Siemens Audiologische Technik GmbH, Erlangen, Germany

Abstract. We propose a novel framework for the personalized design of organic shapes that are constrained to exhibit conformity with the underlying anatomy. Such constrained design is significant for several applications such as the design of implants and prosthetics, which need to be adapted to the anatomy of a patient. In such applications, vaguely defined work instructions are usually employed by expert designers to carry out a sequence of surface modification operations using interactive CAD tools. Our approach involves the abstraction of the work instructions and the expert knowledge into feature dependent machine interpretable rules in a *Knowledge Base*. Robustly identified *canonical* set of anatomical features are then employed to determine concrete surface shaping operations by a *Smart Shape Modeler*. These operations are eventually performed sequentially to adapt a surface to a target shape. The versatility of our approach lies in a priori defining an entire design workflow through a scripting language, thereby yielding a high degree of automation that is completely flexible and customizable via scriptable rules. Consequently, it eliminates tedious manual intervention and offers desirable precision and reproducibility. We validate this framework with a practical application – automatic modeling of shells in hearing aid (HA) manufacturing (HAM).

1 Introduction

Computer aided modeling of organic shapes, such as implants and prosthetics, constitutes an important element of digital manufacturing [5]. Typically, it has two major requirements. First, the designed surfaces should be able to hold essential components such as electronics. Second, they should demonstrate a tolerable degree of conformity with the underlying anatomy. The latter is particularly important to ensure that they comfortably fit to the anatomy of a patient. Due to biological variability, the design of such shapes has traditionally been carried out manually, where expert designers start with a surface representation of the underlying anatomy as an *anatomical template*, which may be acquired through 3D laser or CT scans [5]. The template is then modified by a sequence of manual operations, which are vaguely described in application specific work instructions.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 128–138, 2010.

To this end, the designers rely on certain biomarkers on a template to ensure the conformity between the anatomy and the resulting shape.

The problem is highly significant in various areas, such as HAM, orthopedics, orthotics, and orthodontics. HAM [8], for instance, requires tedious amount of manual involvement. In-the-ear HAs are generally custom made to ensure a comfortable fit to the ear(s) of a patient. A wide range of HA configurations exist to cater for various levels of hearing losses and styles (Fig. 5). HA designers, therefore, have to rely on their experience along with lengthy work instructions for carrying out surface modifications [8] via interactive CAD tools (iCT) [1].

The amount of variation across individuals hinders complete automation. Recently, there have been some advances in this area. In the context of HAs, [6] employed active shape models for describing the shape of a human ear. [7] used Markov random fields for highlighting gender differences. Unal et al. [9] proposed PCA based learning for the design of HA shells. Major limitations of their approach included minimal interpretability for the HA designers, and lack of explicit constraints for component placement. Our approach addresses these limitations by transforming the iCTs to fully automatic tools. In general, iCTs allow a user to specify, for instance, a surface cutting plane by drawing a line in a CAD software, and then to apply the cutting. The design process is simplified if the plane placement and the cut is done automatically. Other examples include automatically painting a region on a surface for local deformation, or positioning additional components at a specific location relative to the surface as in Fig. 1 for HAs.

The proposed framework automates modeling workflows by translating human readable work instructions, and expert knowledge to machine interpretable rules. First, a set of abstract rules is defined in a *knowledge base*, and implemented via a specially designed scripting language. The rules generalize the definition of various selections, e.g., plane definition, and the execution of surface editing operations like cutting. The concrete definition of the operations or selections is determined by a *smart shape modeler* that combines rules with associated anatomical and geometrical features found by an automatic *feature detector*. Once a concrete operation is determined, CAD tools are invoked for actual editing.

Since the rules are implemented via scripts, various surface modeling operations are easily configured for a given application, yielding a high degree of customization and flexibility. The framework automates a process that otherwise exhibits large amount of uncertainty stemming from surface variability. To the best of our knowledge, feature based rules for automated modification of organic surfaces via abstraction of surface modification operations has not been previously proposed. We highlight the effectiveness of our framework by replicating the entire workflow pipeline for the modeling of in-the-ear HAs and substantiate it with experimental results. A fully functional implementation of this framework has already been rolled into the production floor of a major HA manufacturer.

2 Motivation

CAD software allows a user to perform surface shaping operations to adapt them to a final optimal shape. A typical example is provided in Fig. 1, where a sequence

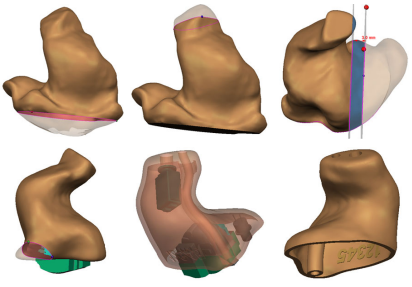


Fig. 1. Modeling operations. Top row: Canal tapering, and helix modification. Bottom row: Vent and electronics placement. The last item corresponds to the finished shell with labeling.

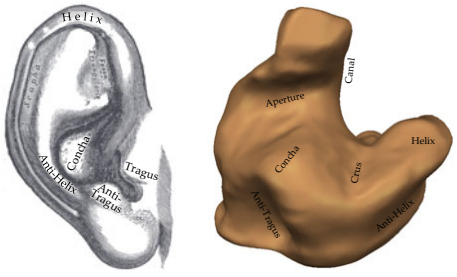


Fig. 2. The human ear anatomy (courtesy [4]) and its 3D reconstruction

of operations is carried out to deform the surface of a 3D outer ear impression to the surface of a finished HA shell. Irrespective of the application, and the nature of an operation, it is typically applied at a particular region or location on a surface. Due to the inherent variability in organic surfaces, it has not been traditionally possible to automate the surface shaping process. A user has to manual specify the operations of interest. Here, we aim for modeling automation by eliminating manual processing to yield consistency, reproducibility, and quality of the designed surfaces. This leads to rapid manufacturing [5], with reduced number of recalls. Although our approach was motivated by HAM, it is fairly general and is applicable to similar applications.

3 Problem Formulation

The design of a surface is defined as a 2-tuple $(\mathcal{M}_s, \mathcal{M}_t)$ and an associated sequence of surface operators $T^{\mathcal{F},\mathcal{O}} = (T_1^{\mathcal{F},\mathcal{O}_1}, \ldots, T_n^{\mathcal{F},\mathcal{O}_n})$, where $\mathcal{M}_s$ and $\mathcal{M}_t$ respectively denote the source and the target shapes:

$$\mathcal{M}_t = (T_n^{\mathcal{F},\mathcal{O}_1} \circ \ldots \circ T_1^{\mathcal{F},\mathcal{O}_n})(\mathcal{M}_s). \tag{1}$$

A given surface is modified when acted upon by a surface operator $\mathcal{M}_i = T_i^{\mathcal{F},\mathcal{O}_i}\mathcal{M}_{i-1}$. Each operator, $T_i^{\mathcal{F},\mathcal{O}_i} : \mathcal{M}_{i-1} \rightarrow \mathcal{M}_i$, is parameterized by a set of anatomical landmarks $\mathcal{F} = \{f_j, j = 1, \ldots, m\}$, and options $\mathcal{O} = \{\mathcal{O}_1, \ldots, \mathcal{O}_n\}$. For simplicity, we drop the explicit dependence of the operator on the *options* for later discussion. Unless specified otherwise, the dependence is always implied.

For a given source, and a pre-specified target shape, $\mathcal{F}$, and $T^{\mathcal{F}}$ are typically defined by process engineers through work instructions, which are executed by trained technicians. Examples of operators include but are not limited to cutting,

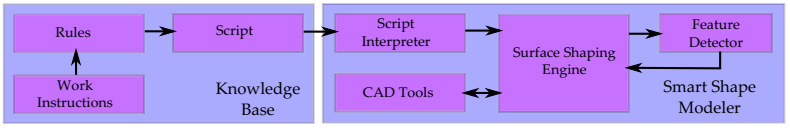


Fig. 3. Smart modeling framework for automation via feature driven rules

rounding, tapering, shrinking, scooping, and even more complex boolean and diffeormorphic operations.

For a given $\mathcal{M}_s$ and pre-specified target $\mathcal{M}_t$, problem under consideration is to estimate the landmarks $\mathcal{F}$ robustly, which will then uniquely identify a subsequence of operations $T^{\mathcal{F}}(R)$ according to certain rules $\mathcal{R} = \{r_k, k = 1, \ldots, K\}$. For complete automation, $\mathcal{F}$ should be consistent for all possible configurations of $\mathcal{M}_s$ and $\mathcal{M}_t$, and rules should be comprehensive to uniquely determine $T^{\mathcal{F}}(\mathcal{R})$. For notational simplicity, we have not indicated the dependence of $\mathcal{F}$ on the surface $\mathcal{M}_{i-1}$, but in general $\mathcal{F}$ does not constitute static landmarks, and have to be determined dynamically.

4 Workflow Automation Framework

Surface shaping steps are generally defined via work instructions. We model this process through a *smart modeling framework* (Fig. 3) that combines this information with the surface landmarks to perform surface shaping operations. Major components include a *knowledge base* (KB), and a *smart shape modeler* (SSM).

4.1 Knowledge Base

The role of the KB is to digitally represent a workflow through a set of rules $\mathcal{R}$ derived from the work instructions. The rules are application specific and encompass all realizations of the class of surfaces. They describe (1) how to perform various steps, and (2) which features to utilize in each step. A step, in turn, consists of one or more surface modification operations $T^{\mathcal{F}}(\mathcal{R})$, and combines features $\mathcal{F}$ for carrying them out. For instance, for cutting a surface, the KB must specify a rule $r \in \mathcal{R}$ that defines where to perform a cut consistently for a wide range of surfaces. This rule should be able to compute the cutting plane, as well as the type of the cut. Although surfaces may exhibit variability, their class membership ensures that some canonical set of features $\mathcal{F}$ is always identifiable, and is sufficient to define such a plane. The existence of the required features drives the entire framework. The KB implements the rules in a scripting language, allowing complete flexibility for various shell models and target shapes.

4.2 Smart Shape Modeler

The function of the SSM is to design a surface (or shape) via CAD tools, which are invoked automatically as directed by the KB. To this end, a *surface shaping engine*

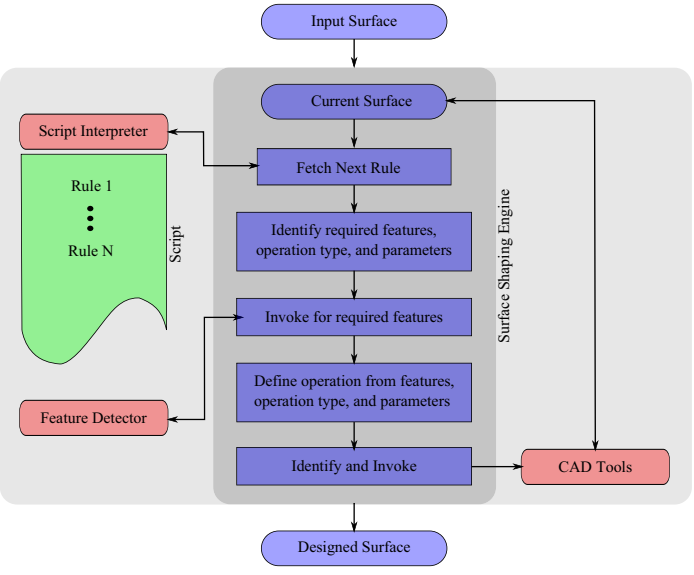


Fig. 4. Interaction of SSE with various components

(SSE) interacts with a *script interpreter* (SI) to sequentially feed it with rules from a digitized workflow. Each rule r is translated to an operation $T_i^F(r)$ that is defined by features $F \subset \mathcal{F}$. Sequential execution of the script rules is ensured through the SI, which parses the rules, and maintains the state of the current rule in a digitized workflow.

As shown in Fig. 4, the role of the SSE is to route the script commands to various components. Once a surface is specified, the SSE resets the current state of the surface, and requests the SI for the next rule. From the rule, it identifies the set of required features, operation type, and the associated parameters (if any). The *feature detector* (FD) is then invoked to detect required features, which are used to uniquely define the actual surface modification operation $T_i^F(r)$. This information leads to a selection of the desired *CAD tool* (CT) with correct parameters to perform T_i on the surface. For instance, for local scooping or smoothing, the KB informs the SSE about (1) the operation (scooping or smoothing); (2) the scooping or smoothing parameters; and (3) the identifier for the corresponding region of interest. Based on this identifier, the FD provides the SSE the area to be scooped or smoothed. The process is repeated until the SI is exhausted of all applicable rules, and the current surface is outputted as the designed shape.

The advantage of this modular approach is that the workflow modeler and the SSE constitute a fully automated smart modeling system that is not specific to a

particular application. One may readily switch between applications by modifying the scriptable rules $\mathcal{R}$ and the associated feature set $\mathcal{F}$. In addition, the smart modeling system is highly configurable via scripting.

5 Application to Hearing Aid Modeling

In this section, we consider the modeling of HA shells [8]. Ear impressions are first acquired by an audiologist by inserting a mold through an injector deep in the ear canal. The mold is allowed to settle, before taking out the impression and carrying out a 3D digital scan. The reconstructed surface then represents the shape of the interior of the ear. During the modeling process, an operator manually shapes an HA and embeds the electronics (Fig. 1). Some major HA configurations are shown in Fig. 5. Despite similar in appearance, ear impressions are unique and exhibit large amount of variability across individuals. The missing data as well as the excess material pose serious challenges to complete automation. Design inaccuracies potentially result in undesirable fit or performance.

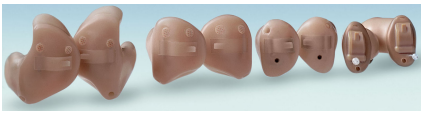


Fig. 5. Various types of HA shells: (Left Most) In-the-Ear (ITE); (Right Most) Completely-In-the-Canal (CIC)

5.1 Feature Detection

A *canonical* set of anatomical features[1] [2] is incorporated in the FD of Fig. 3. It captures comprehensive information about an ear for defining a set of generalized rules. It include points, planes, areas, and curves, and capture the bumps, bends, convexities, concavities, ridges, valleys and the intersections of ridges as illustrated in Figs. 2 and 6. Expert designers already take some of these features into account while manually designing an HA shell. For instance, canal tapering may be done at the second bend plane (Fig. 6(a)), where part of the surface above this plane is removed. Inter-tragal notch (ITN) and crus-side ridge (CSR) (Fig. 6(b)) are used for the placement of vents. Crus area (Fig. 6(c)) is scooped, bumps are rounded, and helix region is tapered for comfort. Feedback seal area is scooped for a firm grip.

We have developed algorithms [2] for the robust detection of these features. The validation of these algorithms was carried out on a dataset of 198 impressions. They were tweaked until a detection rate of 95% and a detection quality in excess of 80% was achieved. Detection quality, in turn, was computed by comparing the detected results with expert manual annotations. In this paper, we assume that these features have already been robustly identified.

[1] These features are consistent across individuals.

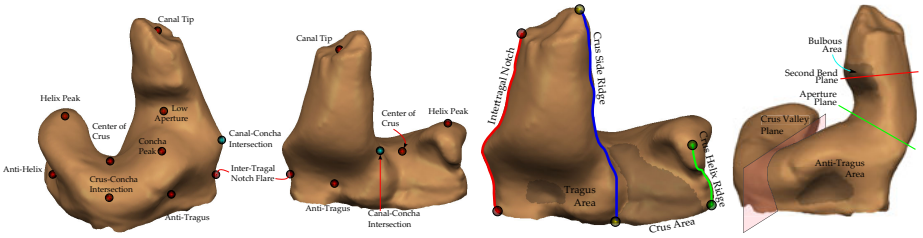


Fig. 6. Anatomical Features: (a) Points; (b) Curves and areas; (c) Planes and areas

5.2 Scripting Language

For realizing the rules in SI, a scripting language with context free grammar is designed using Bison and Flex [3]. The language supports standard data types, in addition to geometric primitives such as points, planes, and matrices, which allow easy manipulation of surface meshes. It includes control structures, such as `if-then-else`, and `for` and `while` loops. Specialized functions include interfaces to CAD tool APIs, FD, and primitive surface manipulators. The reason for developing a customized language is to allow simple handling of the scripts, and its easy integration with a CAD application, while keeping the latter flexible.

5.3 Scriptable Rules

We now briefly describe the role of features in driving rules for performing a typical HA modeling operation. In general, different HA configurations employ different combinations of features, and these combinations are selected through `if-then-else` statements in the script. Examples of rules include `CutCanalAtSeco- ndBend`, `SmoothAtXYZAreaFeature`, and `CutShellWithPlaneDefinedByXYZPoint Features`.

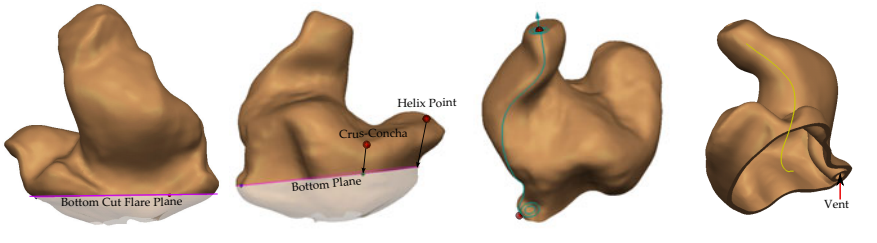


Fig. 7. Automatic placement of the vent: (a) Set up flare plane from anti-helix, tragus, and anti-tragus; (b) Set up the bottom plane by shifting helix peak and crus-concha intersection down and tragus towards the notch; (c) Set up the vent by using the ITN top and bottom points and vent thickness and diameter parameters; (d) Commit vent placement via CAD tool.

Now we describe the usage of the framework, the scripting rules, and the features via the vent placement step. As mentioned earlier, a vent is a tubular attachment to the canal of an HA shell. In the simplest configuration, it is placed along one of the canal ridges because of the acoustic principles and the size constraints. The side selection (ITN or CSR), the starting and ending points, and the wall thickness of a vent may easily be specified through scripting rules. Since the vent runs from the tip of the canal to the bottom of a shell, the script should first define the tip and the bottom planes. The shell bottom may be defined in terms of three point features – helix peak, crus-concha intersection, and tragus.

```
// Move helix peak down by HelixShift
ShiftedHelixPeak = HelixPeak - HelixShift * BottomCutFlarePlane.Normal
// Move tragus
ShiftedTragus = Tragus - TragusShift * BottomCutFlarePlane.Normal
// Move down by ConchaShift
ShiftedCCIPoint = CrusConchaIntersection - ConchaShift * BottomCutFlarePlane.Normal
SanityCheck(ShiftedHelixPeak, ShiftedTragus, ShiftedCCIPoint)
BottomPlane = Plane(ShiftedHelixPeak, ShiftedTragus, ShiftedCCIPoint)
if Below(BottomPlane, Tip) then
    BottomPlane = - BottomPlane
end
```

Once these features are detected, a CAD cutting tool should be invoked on the surface to get rid of the excess material:

```
CutSurface(BottomPlane, Tip)
```

The starting and ending points of the vent are required for its placement. Based on the vent side information, a user may opt to place the vent along ITN, or CSR. The starting and ending points in either case are different, and are found through features, such as the ITN top and bottom points, or CSR top and bottom points. The thickness of the vent wall is taken into account to ensure that the end points fall at the right place on the surface. Eventually, a vent placement CAD tool is invoked to commit the operation:

```
if IsNotchSideVent then
    VentTop = InterTragalNotchTop
    VentDirection = CrusRidgeTop - InterTragalNotchTop
    VentBottom = InterTragalNotchBottom

else
    VentTop = CrusRidgeTop
    VentDirection = InterTragalNotchTop - CrusRidgeTop
    VentBottom = CrusRidgeBottom
end
VentOffset = VentWallThickness + VentDiameter / 2
Normalize(VentDirection)
VentTop = VentTop + VentDirection * VentOffset
PlaceVent(VentTop, TipPlane.Normal, VentBottom, BottomPlane.Normal)
```

KB should contain these instructions in the form of a script, which are sufficient for a primitive vent placement operation. When initialized with a given surface, SSE queries the rule from SI, and executes each instruction. Features are detected on the fly, and the surface modification operation is determined at the run-time. The result is given in Fig. 7. Without loss of generality, these rules correspond to a simplified version of the actual vent placement.

6 Experiments

In this section, we model HAs from 3D ear impressions shown in Fig. 8(a). Results for automatic modeling are given in Fig. 8(b), which indicate excellent canal modification, vent placement, receiver positioning, and labeling even with missing and highly noisy data as in the bottom row.

6.1 Validation

Qualitative Evaluation of Individual Operations. First level of validation was qualitatively carried out by 6 different experts on a dataset of 39 samples. Each expert was asked to rate the outcome of each surface shaping operation according to a *quality matrix* with $\{0 = \text{Unusable}, 1 = \text{Little Modification Required}, 2 = \text{Acceptable}, 3 = \text{Perfect}\}$.

The average of the quality matrix for some common operations as rated by the experts is given in Table 1. Overall the results are very promising. Excellent

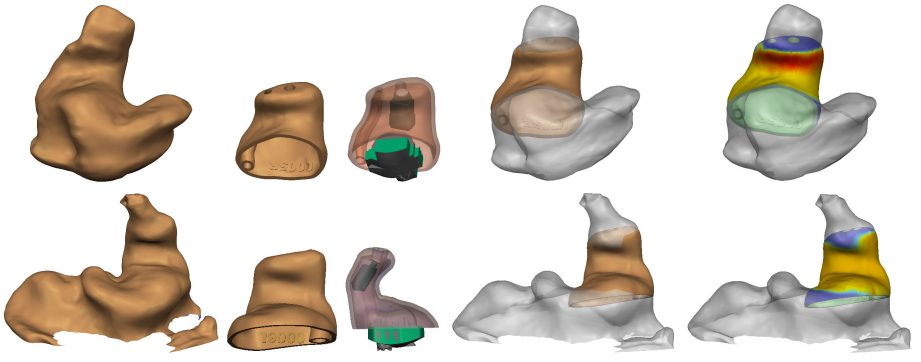


Fig. 8. Automatic detailing: (a) 3D impression; (b) Finished HA shell; (c) Shell with associated electronics; (d) Finished HA shell superimposed on the 3D impression shown with transparency; (e) A 3D error map.

Table 1. Quality matrix for common operations (only for highlighting the appropriateness of the corresponding operations; their description is beyond the scope of the paper)

Operation	Mean ± (Std.Dev.)	Operation	Mean ± (Std.Dev.)
Waxgaurd cut	2.57 (0.86)	ITC crus cut	2.45 (0.78)
Optional vent cuts	2.60 (0.76)	ITE cymba rounding	2.43 (0.77)
Receiver Placement	3.00 (0.00)	Canal Tip cut	1.74 (0.98)
ITE anti-helix filling	2.25 (1.10)	Labeling	1.86 (1.26)
ITC measured cut	2.10 (1.01)	ITE Crus scooping	1.07 (1.14)
Vent placement	2.04 (0.93)	Canal thickening	1.72 (1.13)
Excess material cut	2.21 (0.86)	CIC Measured Cut	1.14 (0.81)
Receiver Hole Placement	1.97 (1.06)	Total	2.08 (0.89)

performance may be noticed for some operations. Although for the others, the experts were not so satisfied, they were still towards the higher end, with the exception of ITE crus scooping, and CIC measured cut. In future, we plan to utilize this matrix as a baseline for fine turning the feature based rules.

Quantitative Evaluation of the Designed Shape. As the second validation, we identified the agreement between a source surface $\mathcal{M}_s$ and the estimated target (designed) surface $\hat{\mathcal{M}}_t$, to quantitatively highlight the fitness of an HA. An error map D was defined on $\hat{\mathcal{M}}_t$ as:

$$D(\mathbf{u}) := \mathcal{S}(\hat{\mathcal{M}}_t(\mathbf{u}), \mathcal{M}_s)||\hat{\mathcal{M}}_t(\mathbf{u}) - \mathcal{M}_s(h(\mathbf{u}))||, \tag{2}$$

where $\mathbf{u}$ parameterizes $\hat{\mathcal{M}}_t$, h represents the closest point mapping from $\hat{\mathcal{M}}_t$ to $\mathcal{M}_s$, and $\mathcal{S}$ denotes a sign function which assigns a positive value if the $\hat{\mathcal{M}}_t(\mathbf{u})$ is outside $\mathcal{M}_s$. The positive values in D indicate where $\hat{\mathcal{M}}_t$ protrudes outside $\mathcal{M}_s$. It should be noted that the source surface represents the inner surface of the ear, and therefore, the positive values means uncomfortable fits. For visualization, we normalize error maps to $[-1, +1]$ prior to applying a colormap with dark blue representing -1, and dark red representing $+1$.

Results given in Fig. 8(e) indicate a high level of agreement with the original 3D impression and the finished product. Red values on the canal correspond to the so-called *feedback seal*. The feedback seal is a selection of a narrow band of triangles on a canal, which is intentionally scooped outwards to ensure a firm grip. It is pointed out that this operation was also performed entirely automatically. Other than the red values corresponding to the feedback seal, mostly yellow and blue values in the error map show that the HA shell stays inside the impression, which amounts to a comfortable fit.

7 Conclusions

This paper has presented a powerful framework for the automation of surface modeling workflows. It translates human readable work instructions to machine interpretable rules. Consequently, interactive CAD operations get simulated by automatic surface shaping operations dictated by the rules. Rules are specified through a special scripting language, and are based on the features detected on a surface. Flexibility of this approach lies in the diversity of rules, which allows handling various configurations in digital manufacturing. The modular nature of the framework makes it fairly general and applicable to a wide range of applications, such as modeling of HAs, dentures, braces, orthotic devices, and orthopedic joint replacements. Results indicate the effectiveness of our approach. With some intervention, it yields industry standard results. At times, a user likes to modify the proposed result, rendering the currently devised rules semi-automatic. In future, we plan to fine tune the rules for full automation, which will allow batching of the work orders.

References

1. 3Shape A/S: ShellDesigner, DentalDesigner, http://www.3shape.com/
2. Baloch, S., Melkisetoglu, R., et al.: Automatic detection of anatomical features on 3D ear impressions for canonical representation. In: MICCAI 2010 (2010)
3. Donnelly, C., et al.: The Bison Manual, Using the YACC Compatible Parser Generator (2003)
4. Henry, G.: Gray's Anatomy: Descriptive and Surgical (1858)
5. Masters, M., et al.: Rapid manufacturing in the hearing industry. In: Hopkinson, N., Hague, R., Dickens, P. (eds.) Rapid Manufacturing: An Industrial Revolution for the Digital Age (2006)
6. Paulsen, R.R., et al.: Building and testing a statistical shape model of the human ear canal. In: Dohi, T., Kikinis, R. (eds.) MICCAI 2002. LNCS, vol. 2489, pp. 373–380. Springer, Heidelberg (2002)
7. Paulsen, R.R., et al.: Shape modelling using Markov random field restoration of point correspondences. In: Taylor, C.J., Noble, J.A. (eds.) IPMI 2003. LNCS, vol. 2732, pp. 1–12. Springer, Heidelberg (2003)
8. Slabaugh, G., Fang, T., et al.: 3D shape modeling for hearing aid design. IEEE Signal Processing Magazine 5(25), 98–102 (2008)
9. Unal, G., et al.: Customized design of hearing aids using statistical shape learning. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 518–526. Springer, Heidelberg (2008)

Manifold Learning for Image-Based Gating of Intravascular Ultrasound(IVUS) Pullback Sequences

Gozde Gul Isguder[1], Gozde Unal[1], Martin Groher[2], Nassir Navab[2], Ali Kemal Kalkan[3], Muzaffer Degertekin[3], Holger Hetterich[4], and Johannes Rieber[4]

[1] Sabanci University,
[2] Technical University Of Munich,
[3] Yeditepe University Hospital,
[4] Ludwig Maximilian University Hospital
{isguder,gozdeunal}@sabanciuniv.edu, {groher,navab}@in.tum.de,
{akalkan,mdegertekin}@yeditepe.edu.tr,
{Holger.Hetterich,Johannes.Rieber}@med.uni-muenchen.de

Abstract. Intravascular Ultrasound(IVUS) is an imaging technology which provides cross-sectional images of internal coronary vessel structures. The IVUS frames are acquired by pulling the catheter back with a motor running at a constant speed. However, during the pullback, some artifacts occur due to the beating heart. These artifacts cause inaccurate measurements for total vessel and lumen volume and limitation for further processing. Elimination of these artifacts are possible with an ECG (electrocardiogram) signal, which determines the time interval corresponding to a particular phase of the cardiac cycle. However, using ECG signal requires a special gating unit, which causes loss of important information about the vessel, and furthermore, ECG gating function may not be available in all clinical systems. To address this problem, we propose an image-based gating technique based on manifold learning. Quantitative tests are performed on 3 different patients, 6 different pullbacks and 24 different vessel cuts. In order to validate our method, the results of our method are compared to those of ECG-Gating method.

Keywords: Manifold Learning, Classification, IVUS, Image-based gating, ECG gating.

1 Introduction

Intravascular Ultrasound is a unique invasive catheter-based imaging technology, which yields a high resolution, real-time cross-sectional view of the blood vessels from the inside-out. The cross-sectional images are acquired by pulling the catheter back with a motor running at a previously defined constant speed, and this whole process is referred as a pullback. Since IVUS modality provides a very detailed information about the internal vessel structures, it is a unique

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 139–148, 2010.

tool for the diagnostics of coronary artery diseases(CAD) and plaque characterization. For diagnosis and assessment of the disease, accurate measurements of the total vessel and the lumen volume in the suspicious lesion areas are crucial. However, quality of the IVUS evaluations, and accuracy of the measurements deteriorate due to artifacts caused by heart movement during a pullback[1]. The most obvious artifact is the back and forth movement of the catheter in the vessel longitudinal direction due to the periodical change in the blood flow while the heart muscles are contracting and expanding. In [2], the authors observe that the IVUS transducers within coronary vessels have a longitudinal movement of average 1.50 ± 0.80 mm during each cardiac cycle. As the transducer moves back and forth, it passes through the same locations of the vessel multiple times; thus it oversamples the vessel. This means gaining unnecessary information which leads to computational inefficiency for further processing. Furthermore, due to the movement, the longitudinal cut of the vessel has a saw-toothed appearance(see Fig. 5 first row) which makes the segmentation of the vessel even harder. Another artifact caused by the cardiac cycle is the change of the vessel morphology due to the varying blood pressure during the cycle. The change in the morphology leads to the variations in the lumen area observed at different cardiac phases(systole,diastole). In [1,2], it is stated that measured lumen and vessel volumes in non-gated image sets are significantly larger than normal and the choice of the suitable phase is still a question. A way to account for the problems above is introducing an electrocardiogram (ECG) signal, which is capable of giving information about the heart's current physical status. By utilising the ECG signal, heart and IVUS transducer are synchronized so as to capture the frames only near the predetermined fraction of the RR-interval[1]. However online-ECG gating requires an ECG unit, which may not be always available to the physician. Furthermore, in the older systems that used ECG triggering, ECG gating increased the image acquisition time, and in the new systems, the acquisition time is not affected but some important information about the vessel is lost.

In this paper, we introduce a robust image-based gating method based on manifold learning. By designing this method, our overall aim is to retain only the necessary information about the vessel, (the frames at a particular fraction of the RR-interval), which will be good enough to provide accurate lumen volume measurements and vessel length; at the same time will avoid loss of important plaque information in the lesion areas.

2 Related Work

In [3], a method for classification of the IVUS frames as end-diastolic and not end-diastolic is presented. As preprocessing, important image characteristics such as edges are enhanced, different feature vectors based on spatial and frequency characteristics of the images are defined, and finally a nearest neighbor search based on the Euclidean distance between the feature vectors is used to classify the frames.

In [4], a method to retrieve the cardiac phase by examining the features in a circular region of interest, namely Average Intensity(AI) and Absolute Intensity Difference between the subsequent frames, was discussed.

In [5,6], an image-based gating algorithm is proposed where a Dissimilarity Matrix based on the Normalized Cross Correlation (NCC) is built between each 2-tuples of the pullback; then the matrix is filtered with an X-shaped inverted Gaussian kernel, which highlights the frames with high similarity. Finally an algorithm to find the highest local maxima along the path on the diagonals, that represents the optimal gating frames, is introduced.

In [7], the authors use a similar technique to [5], by building the dissimilarity matrix D based on the image descriptors which are defined based on Gabor patches. A 1D signal is extracted from D, which defines the similarities between the frames and finally a local minimum search over the 1D signal is performed to obtain the best frames.

3 Our Contribution

In all the methods discussed above, even if different techniques were used, the overall objective is to be able to construct a 1D signal similar to R-waves by using the information(features) that is embedded in the images. In this paper, we propose directly projecting our high dimensional data, to a low-dimensional manifold and thus treating each image frame as a low-D signal in the low-D manifold.

A variety of dimensionality reduction techniques have been proposed in the literature, ever since emergence of complex and high-dimensional input data. The most commonly used linear dimensionality reduction techniques such as Principal Component Analysis(PCA) and Multi-dimensional scaling(MDS) are efficient and simple, however are not able to detect nonlinear structures that exist almost in all true datasets. The human cardiac system is nonstationary,dynamic and nonlinear; hence, linear analyses may not account for all aspects of cardiac performance[8,9].

Manifold learning is an effective, geometrically motivated, nonlinear dimensionality reduction technique, which is used to solve a variety of vision problems such as segmentation, registration, tracking and object recognition. The technique was validated to be successful, particularly if the input has smooth appearance variation or smooth deformation[12]. As explained in Section 1, cardiac cycle's first effect, slowly varying longitudinal movement of the catheter, results in a smooth appearance variation. Furthermore, the slight vessel morphology change during the cycle results in a smooth deformation in the input images.

In addition, in the lesion areas where the cross-sectional view of the vessel can change faster, using global distance metrics would fail because the lesion areas would be detected as outliers. However manifold learning preserves locality, which makes it much less sensitive to noise and outliers.

With the motivation provided above, our contribution in this paper, is to adopt and apply the manifold learning framework to our problem of image-based gating of IVUS pullbacks as explained in the next section.

4 Method

Isomap [10], local linear embedding [11], and Laplacian eigenmaps [12] are three different techniques for manifold learning. In this paper we will use Laplacian eigenmaps technique, which is very simple and efficient since it solves only one sparse eigenvalue problem. We construct our problem as follows:

Given a set of k points $x_1, ..., x_k$ in $\mathbb{R}^d$, where k is the number of frames x_i in the pullback and d is the dimension of the image; find another set of k points $y_1, ..., y_k$ in $\mathbb{R}^m$, where $m \ll d$. We assume that $x_1, ..., x_k \in \mathcal{M}$, where $\mathcal{M}$ is a manifold embedded in $\mathbb{R}^d$.

An important issue is the choice of the dimension of the $\mathcal{M}$, denoted by m. We need one dimension to account for the smooth appearance variation caused by

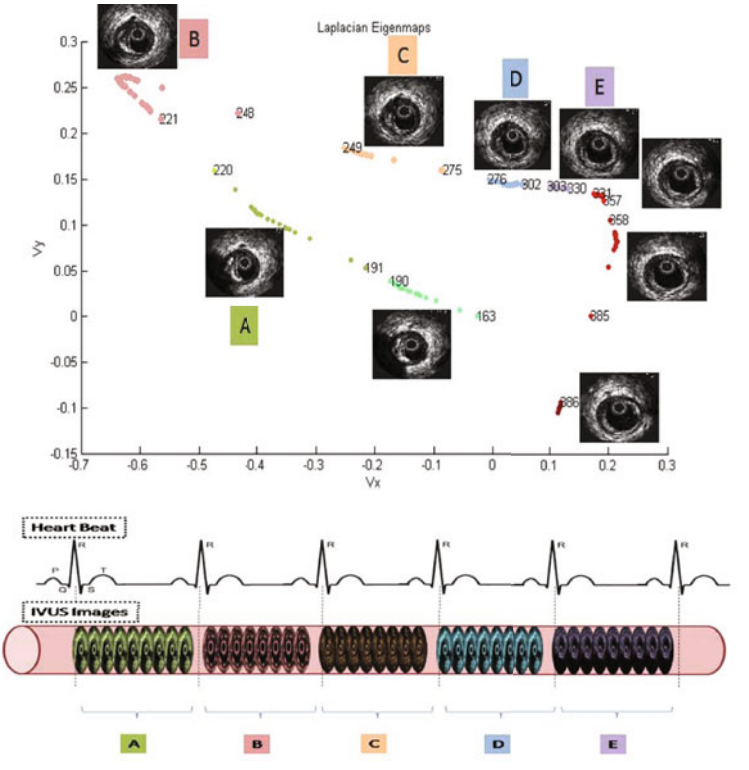


Fig. 1. An illustration of Manifold idea. Each frame is shown with a dot on the calculated low-D manifold(here m=2), where A,B,C,D,E are the clusters of frames that belong to different cardiac cycles.

the first artifact, and another dimension to account for the smooth deformation caused by the second artifact (see Section 3). Thereby, we choose $m \geq 2$. We heuristically choose $m = 3$. In another words we choose to represent each image frame with a 3D vector.

4.1 Laplacian Eigenmaps

Laplacian eigenmaps[12] is an approach that incorporates the neighborhood information of the input to build a weighted graph. After building the graph, a low-D representation of the input that optimally preserves local neighborhood information, is computed by using the Laplacian of the graph. It is worth noting that, since the approach is geometrically motivated, the resulting mapping will be a discrete approximation of a continuous map from the high dimensional space to the low-D manifold.

The first step is to construct the graph with representative k nodes for each x_i and edges between the nodes x_i and x_j , if the nodes are close enough. The relationship of being close can be defined as an ϵ neighborhood $||x_i - x_j|| < \epsilon$, where $||.||$ denotes the Euclidean norm. The disadvantage of this choice is the parameter setting. Another option is using n nearest neighbors, where one can put an edge between the nodes x_i and x_j if j is one of the n nearest neighbors of i. If we define the approximate number of frames in each cardiac cycle as n_{cycle}, the parameter n should satisfy $n \geq n_{cycle}$.

Another important issue is defining a similarity measure for the nodes x_i and x_j. In our weighting function we used Sum of Squared Distances(SSD), but different measures such as Sum of Absolute Difference(SAD) or Normalized Cross Correlation (NCC) can also be used.

The second step is to weight the edges in the graph by the appropriate weights. Weight function is inspired from the heat equation given by

$$W_{ij} = exp^{-||x_i - x_j||^2/2\sigma^2} \tag{1}$$

where σ^2 is the variance. $W = [W_{ij}]; ij \in [1, .., k]$, forms the weight matrix.

As a final step, a diagonal weight matrix D is constructed by summing up the coloumns of W, and Laplacian $L = D - W$ is then calculated. The eigenvectors corresponding to the smallest eigenvalues (excluding zero) of the Laplacian matrix gives the desired mapping. We refer to [12] for further details on the Laplacian eigenmap method.

4.2 Clustering

In general there are about ≈ 90 cardiac cycles in each pullback(see Table 2). However, to illustrate the idea of manifold learning, a desired mapping for 9 sequential cardiac cycles is given in Fig. 1. In Fig. 1, the low-D manifold gives a very nice intuition of the clusters that belongs to the same cardiac cycle. We observed that the largest distance between the eigenvectors V_i and V_{i+1} occurs if V_i is the last frame that belongs to the n^{th} cardiac cycle and V_{i+1} is the starting

frame of the $(n+1)^{st}$ cardiac cycle. The distance signal $d = ||V_{i+1} - V_i||$ is constructed. Since the method preserves the local-distances, the global structure is more visible after normalization of d. In Fig. 2, constructed d function for 400 frames is shown for illustration, where green boxes indicate the local maxima points in d function.

In order to find the local maxima points of the distance function d, we utilise a morphological eroding operation. Let $min_{cluster}$ be the minimum number of frames in a cluster, then a structuring element of size $min_{cluster}$ is constructed and distance signal is eroded with the structuring element. If the current point is a local maximum around the structuring element, then eroding operation will not change its value, thus we check for the points that has not changed after the eroding operation. This very simple but efficient technique finds the local maxima points. In some cases, where the frames of the sequential cardiac cycles are too similar, e.g branching areas, the distance function may not have any local maxima. In other cases, where the frames of one cardiac cycle are too different (e.g lesion areas where vessel changes rapidly), the distance function may have more than one local maxima. In those cases, we refine the results of the local maxima and check the number of frames between each possible consecutive maximum. If the difference between the two consecutive possible maximum points is bigger than $2n_{cycle}$, we look for another local maximum between them, and if the difference is smaller than $n_{cycle}/2$ we eliminate the possible maximum. After the post-processing step, local maxima of the distance function d, hence the gated(stable) image frames are obtained, and the IVUS gating algorithm is completed.

5 Experiments and Results

We applied our automatic image-based gating algorithm on 3 different patients and 6 different pullbacks. All the pullbacks were acquired in-vivo in the coronary

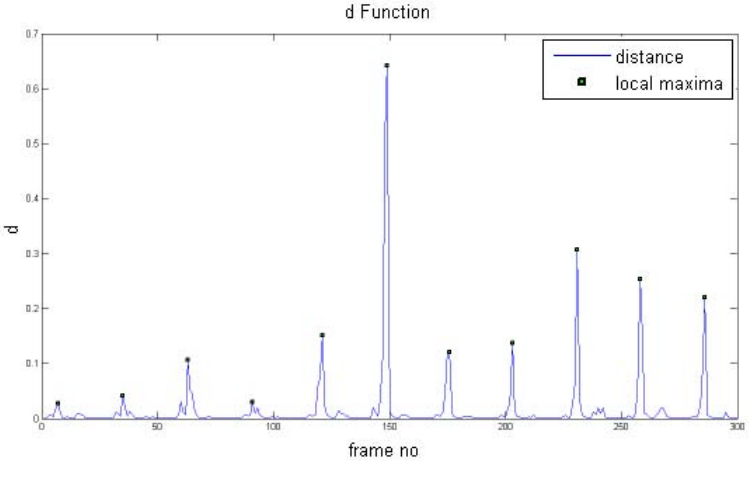


Fig. 2. Normalized Distance Function

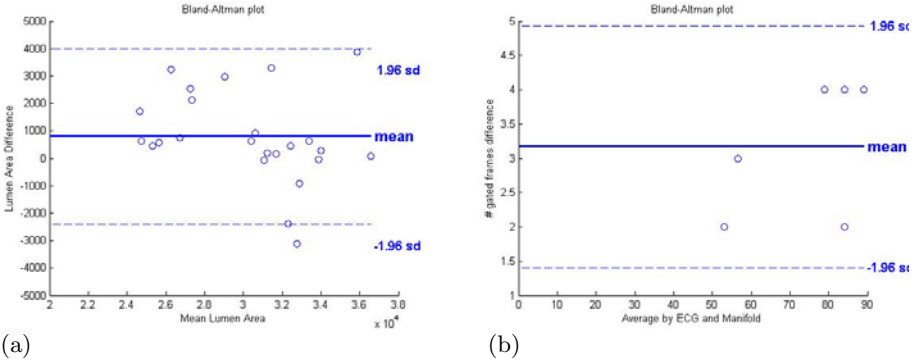
Table 1. Lumen Area Differences Error Analysis, where pt_{id} is the patient id, pb_{id} is the pullback id, angle is the viewing angle for constructing a longitudinal cut and LAD Error is the lumen area difference error

pt_{id}	1	2	2	3	3	3
pb_{id}	1	1	2	3	1	2
Angle	10 50 130 150	10 50 130 150	10 50 130 150	10 50 130 150	10 50 130 150	10 50 130 150
LAD Error	0.02 0.04 0.02 0.02	0 0.05 0.1 0.1	0.07 0.03 0.07 0.09	0.02 0 0.03 0	0.03 0.01 0 0.02	0.08 0.1 0.01 0.01
Mean Error	0.02	0.06	0.06	0.01	0.01	0.05

arteries of the patients of our clinical partners with 40 MHz IVUS catheter. The frame rate was 30 Hz and the motorized pullback speed was 0.5 mm/s.

In our method, we used n_{cycle} to represent the number of frames per cardiac cycle. n_{cycle} is defined as $f_{Rate}/1.2Hz$, where 1.2 Hz is the average heart beat rate of human species and f_{Rate} is the frame rate of the pullback. Similarly $min_{cluster}$ is used to represent the minimum number of frames in a cardiac cycle. $min_{cluster}$ can be equated to $f_{Rate}/(1.2Hz + 2\sigma)$, where σ is the variance of the heart beat rate. The variance of the heartbeat may be high in the patients with irregular heart beats. Choosing a large σ would guarantee to find the local maxima points in d function for those patients. However, the latter may lead to too many possible local maxima points. In our experiments, we choose a small $\sigma = 0.1Hz$.

In order to validate our results, we compared the number of gated frames obtained by our algorithm and by ECG gating algorithm. In Table 2, the number of gated frames from the two methods show agreement. As stated in Section 3, the accurate measurements of lumen is crucial for coronary artery diseases' diagnostics. For that reason, we compared the lumen areas calculated from our gated pullbacks and ecg gated pullbacks. The lumen areas were drawn by the expert cardiologists in our team. For more accurate results, we compared the



(a) (b)

Fig. 3. (a)Bland-Altmann Analysis of Lumen Areas drawn by the medical experts on ecg gated pullback and image-based gated pullback: 790±40.79 pix. (b) Bland-Altmann Analysis of gated frame normalized count calculated by ecg gating and image-based gating: 3.1667 ± 0.937.

Table 2. Comparison of the frame counts chosen by ecg gating algorithm and our image based gating algorithm.pt_{id} is the patient id, pb_{id} is the pullback id, $\#ecg$ represents the number of frames chosen by ecg, $\#alg$ shows the frame count selected by our algorithm and length[mm] is the actual length of the vessel.

pt_{id}	pb_{id}	#total	$\#ecg$	$\#alg$	length[mm]
1	1	2328	81	77	38.80
2	1	1627	54	52	27.12
2	2	1800	58	55	30.00
3	1	2388	85	83	39.80
3	2	2379	86	82	39.65
3	3	2358	91	82	39.30

lumen areas in the longitudinal views of the vessel from 4 different angles for each pullback. In Fig 4, an illustration of vessel longitudinal cuts at different angles is given. Difference(LAD) error in Table 1 is calculated as the ratio of the absolute difference of the areas found by the two methods, with the ecg-gated pullback area used as the ground truth: $LAD_{error} = |LA_{ecg} - LA_{alg}|/LA_{ecg}$, where LA_{ecg} is the lumen area in the ecg gated pullback and LA_{alg} in the image based gated pullback. An LAD error rate of 0.04±0.03 is obtained. In addition, a Bland-Altmann analysis on the lumen areas (Fig. 3.a), revealed that more than 95% of the measurements were in agreement between the two methods. In Fig. 3.b, a plot for the Bland-Altmann analysis based on the number of gated frames calculated by both methods is given. To account for the different vessel lengths among our dataset, we considered the normalized counts, computed as #gated / #total. Longitudinal IVUS views shown in Fig 5, demonstrate similar qualitative outcome for our manifold-learning based IVUS gating method, and the ECG-gated method, which is also verified by the expert cardiologist in the team.

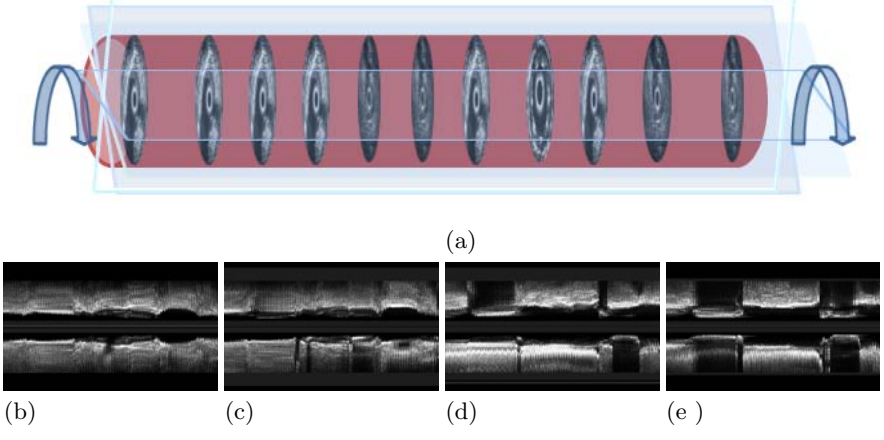


Fig. 4. (a)An illustration of longitudinal cuts(LC) at different angles (b) LC from 10° (c) LC from 50° (d) LC from 130° (e) LC from 150°.

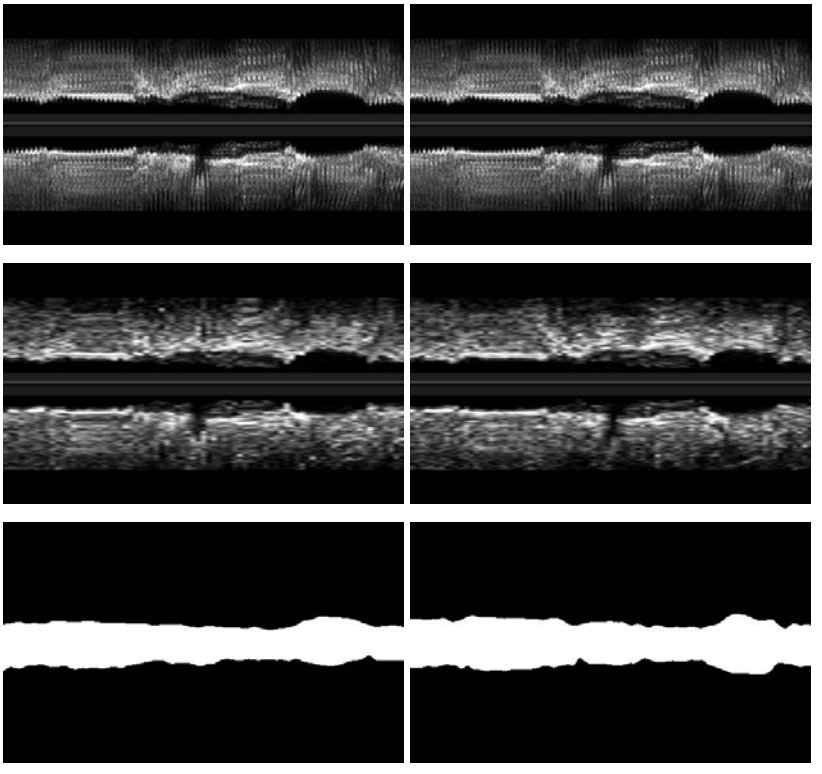


Fig. 5. *First row:* Nongated pullback. *Middle Row; Left:* Image-based gated pullback. *Right:* Ecg gated pullback. *Bottom Row; Left:* Lumen Area of image based gated pullback. *Right:* Lumen Area of ecg gated pullback.

6 Conclusion

We presented a novel image-based gating method for IVUS sequences. Our method is based on manifold learning, which embeds the similar IVUS frames onto contiguous positions of a low-dimensional manifold lying on a high dimensional image space. Further, we classified the frames by using distances between consecutive eigenvectors that represent the IVUS frames using the frame rate of the pullback and basic heart beat rate knowledge. We tested our data on 3 patients and 6 in-vivo pullbacks. We compared the number of selected frames and the lumen areas in 4 different longitudinal views, computed by both methods. Future directions for this work include analysis of lumen volume differences and the plaque areas of the gated pullbacks.

References

1. Bruining, N., von Birgelen, C., de Feyter, P.J., Ligthart, J., Li, W., Serruys, P.W., Roelandt, J.: Ecg-gated versus nongated three-dimensional intracoronary ultrasound analysis: Implications for volumetric measurements. Catheterization and Cardiovascular Diagnosis 43, 254–260 (1998)
2. Arbab-Zadeh, A., DeMaria, A.N., Penny, W., Russo, R., Kimura, B., Bhargava, V.: Axial movement of the intravascular ultrasound probe during the cardiac cycle: Implications for three-dimensional reconstruction and measurements of coronary dimensions. American Heart Journal 138, 865–872 (1999)
3. de Winter, S.A., Hamers, R., Degertekin, M., Tanabe, K., Lemos, P.A., Serruys, P.W., Roelandt, J.R.T.C., Bruining, N.: A novel retrospective gating method for intracoronary ultrasound images based on image properties. Computers in Cardiology, 13–16 (2003)
4. Zhu, H., Oakeson, K.D., Friedman, M.H.: Retrieval of cardiac phase from IVUS sequences. In: Medical Imaging 2003: Ultrasonic Imaging and Signal Processing, vol. 5035, pp. 135–146 (2003)
5. O'Malley, S.M., Carlier, S.G., Naghavi, M., Kakadiaris, I.A.: Image-based frame gating of IVUS pullbacks: A surrogate for ecg. In: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2007, April 15-20, vol. 1, pp. I-433–I-436 (2007)
6. O'Malley, S.M., Carlier, S.G., Naghavi, M., Kakadiaris, I.A.: Image-based frame gating for contrast-enhanced IVUS sequences. In: Proceedings of Computer Vision for Intravascular and Intracardiac Imaging (CVII), Copenhagen, Denmark (October 2006)
7. Gatta, C., Pujol, O., Leor, O.R., Ferre, J.M., Radeva, P.: Robust image-based ivus pullbacks gating. Med. Image. Comput. Assist. Interv. Int. 11(Pt 2), 518–525 (2008)
8. Schumacher, A.: Linear and Nonlinear Approaches to the Analysis of R-R Interval Variability. Biological Research For Nursing 5(3), 211–221 (2004)
9. Malpas, S.C.: Neural influences on cardiovascular variability: possibilities and pitfalls. Am. J. Physiol. Heart. Circ. Physiol. 282(1), H6–H20 (2002)
10. Tenenbaum, J.B., Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science 290(5500), 2319 (2000)
11. Roweis, S.T., Saul, L.K.: Nonlinear Dimensionality Reduction by Locally Linear Embedding. Science 290(5500), 2323–2326 (2000)
12. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput. 15(6) (2003)

Automatic Computation of Electrodes Trajectory for Deep Brain Stimulation

Caroline Essert[1,2,3,4], Claire Haegelen[2,3,4,5], and Pierre Jannin[2,3,4]

[1] LSIIT - University of Strasbourg, Pôle API, Boulevard S. Brant, Illkirch, France
[2] INSERM, U746, Faculty of Medicine, Rennes, France
[3] INRIA, Centre Rennes - Bretagne Atlantique, VisAGeS Unit/Project, Rennes, France
[4] University of Rennes I, CNRS, UMR 6074, IRISA, Rennes, France
[5] Department of Neurosurgery, Pontchaillou University Hospital, Rennes, France
essert@unistra.fr

Abstract. In this paper, we propose an approach to find the optimal position of an electrode, for assisting surgeons in planning Deep Brain Stimulation. We first show how we formalized the rules governing this surgical procedure into geometric constraints. Then we explain our method, using a formal geometric solver, and a template built from 15 MRIs, used to propose a space of possible solutions and the optimal one. We show our results for the retrospective study on 8 implantations from 4 patients, and compare them with the trajectory of the electrode that was actually implanted. The results show a slight difference with the reference trajectories, with a better evaluation for our proposition.

1 Introduction

Nowadays, an increasing number of patients suffering from Parkinson's disease or essential tremors are treated by Deep Brain Stimulation (DBS). This intervention consists in implanting an electrode in a deep location of the brain, in order to stimulate a zone with an electric current, causing an inhibition of the disease effects. This treatment is very efficient, but also very difficult to plan. The tedious planning phase, mainly relies on the study of the patient images (such as MRI and CT), acquired before the intervention. Sometimes, a safe planning can not be found, prohibiting such an intervention for the patient. The objective of the work presented in this paper is to provide the neurosurgeon with a planning tool able to assist him in finding the optimal linear trajectory for a DBS electrode.

In the domain of assistance to surgical planning, various tools already exist, for example to help in finding automatically the target [1]. However, we focus in this study on the placement of the surgical tools. In surgical planning in general, a lot of planning tools are simulators allowing to model what will be the effect of a treatment, for a given placement of electrodes proposed by the surgeon. However, this forces the surgeon to perform himself/herself the trial and error search which may be a tedious task. Some authors proposed interesting attempts of automatic targeting methods, for various kinds of surgeries. However, they also have some drawbacks that we would like to overcome. In [2], which focuses on hepatic RFA needle placement, authors do not take into account the presence of surrounding organs. Authors of [3] and [4] confess a long computation

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 149–158, 2010.

time, and the first one's algorithm is only in 2D. In [5], focused on DBS electrodes placement, authors restrict the search to a limited set of possible entry points, avoiding possibly good trajectories to be discovered. We prefer to let the software decide by itself possible entry points, in order to study all possible solutions as long as they satisfy the rules of the intervention.

In this paper we present a method based on the resolution of geometric constraints, inspired from [6,7], to automatically compute a safe path to the target for DBS planning. We chose to develop a generic approach, but to restrict our experimental study and validation to the targeting of the Sub-Thalamic Nucleus (STN). Our method is based on 2 types of data: the pre-operative patient-specific images, and the rules specific to each type of intervention. We first detail in Section 2.1 the approach we used to determine the rules of DBS planning, and detail how we treated them, either formalizing them into geometric constraints or simply adjusting some parameters. Then in Section 2.2, we explain how we obtained the patient-specific data from images. In Section 2.3, we expose our approach and the formal solver we developed to solve geometric constraints with image data. Section 2.4 exposes the formalized constraints we defined for DBS planning. Finally, we summarize our results on the retrospective study of 8 implantations from 4 patients, and discuss the comparison between our results and the reference trajectory segmented from post operative CT images.

2 Materials and Methods

2.1 Analysis of Rules Governing DBS Planning

First, we examined the literature about DBS planning and intervention procedure, in order to have a first idea of the overall planning process and of the main rules used by the neurosurgeons when selecting an optimal path. This enabled us to prepare a set of questions for the 2 neurosurgeons who participated in our study: one of them with an experience of approximately 100 DBS implantations, and the other with about 200 implantations. We had 2 interviews with these neurosurgeons. Both interviews also allowed us to discover new rules, and to progressively refine our set of rules. We also went to the operating room to observe planning processes and interventions, and had a third interview, in order to precisely define the main rules to use in our study. The rules we chose and their formalization or parameterization are presented below.

1. **Placing the Electrode in the Target.** Our first obvious rule is that the electrode tip must be in the target (in our case the STN). This rule is necessary to restrict the field of research of a position. Our solver expects the definition of a target, and has been natively designed to consider only 3D positions of the tool having the tip inside the target as possible solutions.
2. **Position of the insertion point.** The insertion point on the head has to be in the upper surface of the skin of the head, as the surgeon will never implant the electrode through the lower parts of the head. This is not only due to accessibility reasons, but also to aesthetics reasons. We provide our solver with an initial insertion zone corresponding to the scalp of the head.

3. **Path length restriction.** This rule concerns the maximal length of the path, which restricts the field of research. According to neurosurgeons instructions, we formalized this rule by assuming that the path has to be shorter than 90mm.
4. **Avoiding Risky Structures.** The third rule is to find an electrode placement that avoids crossing vital or risky structures. For DBS, the identified "obstacle" structures include the ventricles and the vessels. Vessels are numerous in the brain, and usually located in the of the cortical sulci. Unfortunately they are often invisible when images are acquired without contrast agent or without angiography. So the neurosurgeons usually rely on the anatomical location of the sulci and avoid trajectories passing through them. We also choose to avoid the bottom of the sulci to be conform to their methods, and to be generic enough in case no segmentation of the vessels is possible. We formalized this rule by assuming that the insertion point in the skull of the patient has to be visible from the target, without any occlusion by one of the cerebral structures considered as obstacles.
5. **Minimizing the length of the path.** Even if we are sure that the path is shorter than the maximum length defined by rule # 3, minimizing the length of the path as much as possible reduces the risks of a bending of the electrode. We formalized this rule by assuming that the proportion between the length of the path and the shortest distance between target and the skin mesh has to be minimal.
6. **Maximizing the Distance Between Electrode and Risky Structures.** Even if we are sure that the electrode will not cross any risky structure thanks to rule # 4, it is less risky if the trajectory passes as far as possible from those structures. We formalized this rule by assuming that the distance between the electrode and the structures designated as risky has to be maximized.
7. **Optimizing the Orientation of the Electrode According to Target Shape.** The surgeons also expressed their wish, when selecting a trajectory, to have its axis close to the main axis of the target. This way, they can try several possible depths to cover almost all of the target without needing another insertion at a different location. We formalized this rule by assuming that the angle between the axis of the electrode and the main axis of the target has to be minimized.

2.2 Data

The proposed algorithm needs to be provided with a set of spatially defined objects, such as: an initial insertion zone, the target, and all structures that come into play in the constraints as obstacles or risky structures. In this section, we describe how we obtained those data.

All patients we retrospectively studied had the same image acquisitions: one pre-operative 3-T T1-weighted MR (1 mm x 1 mm x 1 mm, Philips Medical Systems), one pre- and one post-operative CT scans (0.44 mm x 0.44 mm x 0.6 mm in post-operative acquisitions and 0.5 mm x 0.5 mm x 0.6 mm in pre-operative acquisitions, GE Healthcare VCT 64). The MRI and the pre-operative CT were acquired just before the intervention, and the post-operative CT was acquired 1 month after. The three images were denoised with a Non-local means algorithm. A bias correction algorithm based on intensity values was also applied on MR images. CT images were rigidly registered to pre operative MR images, using the Newuoa optimization. From these co-registered

images, we segmented the ventricles (from MRI), the cortical sulci (from MRI), and the reference electrode for validation (from post-operative CT scans).

We also used a mono-subject anatomical template, made from 15 3T MR acquisitions at high-resolution [8], to segment the initial zone once for all. Thanks to an affine registration, this surface mesh was adapted to each studied patient's MRI. Finally, anatomical structures and objects were defined in the same common space, allowing us to compute the trajectory constraints.

In this study, we focused on the computation of the trajectory to reach a target, and not on the delineation of this target. Our goal was to compare our proposition with reference trajectories. To this end, we needed to use the exact same target position as the one actually performed during the intervention. That is why we chose to segment the contacts of the electrode in the post-operative CT and used them as the target.

2.3 Global Strategy

Our strategy consists in formalizing the rules described in Section 2.1 into geometric constraints in order to solve them with a constraints solver. Among those constraints, some are boolean (#1, 2, 3, 4) and others are numerical (#5, 6, 7). The first ones are called *strict constraints*. They have to be satisfied necessarily, and define the space of possible solutions. The second ones are *soft constraints* that need to be optimized at best, according to a weighting factor defined by the surgeon. Among the space of possible solutions, the optimal path will be the one that satisfies the soft constraints at best. Let us note that constraints #1 and 2 are not solved but included as input image data in our solver (target and scalp).

We developed our own geometric constraints solver in C++, based on MITK software system, and using ITK and VTK libraries. We gave a great importance to the genericity in our approach, our goal being to dispose of a generic solver able to be used for any surgical intervention involving a path planning for a rectilinear tool. Therefore, our solver takes as input data not only the images and segmented cerebral structures, but also the rules of the intervention written in a specifically defined meta-language, and written in a separate XML file which is loaded when the software is launched. If we want to add an extra constraint, we only have to write it in this file.

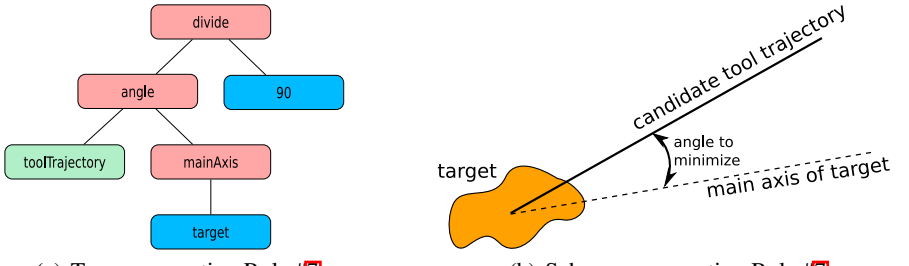
A solution is constituted by a position in the 3D space for the electrode. It can be represented indifferently either by a point (*i.e.* the tip of the electrode) and a direction, or by two points (coordinates in $\mathbb{R}^5$ are sufficient). We chose to use the second alternative that was more intuitive, by using the tip and the insertion point on the skin. We start with an *initial solution space* constituted by the mesh of the *initial insertion zone* constituted by the scalp of the skin for the insertion point, and the whole target for the tip point.

The resolution is performed in two steps. The first phase consists in reducing the initial insertion zone by eliminating the triangles of the mesh that do not satisfy the totality of the strict boolean constraints. The second phase consists in a numerical optimization of the soft numerical constraints. Each soft constraint corresponds to a cost function to minimize. In order to take into account all constraints with a weighting factor defined by the physician, we chose to combine them into an aggregative cost function. After an initialization of the process, consisting in a rough evaluation of the values at some insertion points homogeneously spread over the zone of possible insertion points, we

compute some connected components around the best candidate points, and we start an optimization using Nelder-Mead optimization method from the best candidate. This way we avoided to fall into local minima, as we proved in a previous paper [7].

2.4 Geometric Constraints

Using our meta-language, the rules (or their corresponding cost function) are translated as geometric constraints represented as terms, combining operators and known data, according to a geometric universe. The geometric universe we defined for our constants and unknowns includes the usual types (*e.g.* integers, real numbers, booleans) and composed types such as *point*, *tool*, *shape*, or *solution*. We also defined a certain number of operators: usual operators such as plus, minus, multiply, and, or, as well as complex operators as for instance *distMin*, *distToolOrgan*, *angle*, *visible*. In order to add an extra constraint in the XML file, the necessary operators must have already been defined. The terms can be seen as trees, which are solved using a depth-first approach.



(a) Tree representing Rule #7. (b) Scheme representing Rule #7.

Fig. 1. Different representations of Rule #7.

As an example, let us analyze Rule #7. This rule aims at optimizing the orientation of the electrode according to the shape of the target (as shown on Fig. 1(b)). It is translated into a soft geometric constraint expressing that the angle between the trajectory of the electrode and the main axis of the target has to be minimal. It is computed by the minimization of a numerical cost function $f_{orientation} : \mathbb{R}^5 \rightarrow [0, 1]$. This cost function is chosen in a way that the resulting values are between 0 and 1, in order to obtain an order of magnitude comparable to the cost functions of the other rules before combining them. Without this normalization, a rough combination of these functions would be meaningless. So we transform the formalization by saying that the ratio between the angle and 90 has to be minimal. This way, $f_{orientation}$ tends to 0 if the angle is close to 0, and to 1 if the angle is close to 90 degrees. Function $f_{orientation}$ is then defined by (1). In this equation, X represents the degrees of freedom in $\mathbb{R}^5$ of the trajectory of the tool.

$$f_{orientation}(X) = \frac{angle\_between(X, axis\_target)}{90} \tag{1}$$

To express this function as a constraint understandable by our solver, we use our meta-language and write it as a term. This term uses: existing operators defined in the solver (*divide*, *angle*, *mainAxis*), constant data (target shape coming from the images, integer 90), and the variable which will be filled with a candidate value (*toolTrajectory*). The final constraint in XML syntax is shown in Table 1. The corresponding tree is illustrated on Fig.1(a): operators are in red, given constant data are in blue, and the variable is in green. In the solver, we use this tree structure to represent the constraints. If a data or variable node is used in more than one constraint, it exists only once and doesn't have to be re-evaluated several times.

Table 1. XML formalization of Rule #7

```
<soft_constraint name="optimized orientation" label="sc_ori">
divide( angle ( toolTrajectory, mainAxis ( target ) ), 90.0 )
</soft_constraint>
```

This way, all the constraints we detailed in Section 2.1 were written in XML syntax using our operators and data. One last soft constraint is added, representing the aggregative constraint which combines the previously defined soft constraints with the chosen weighting factors, and corresponds to the aggregative cost function f_{final} (2).

$$f_{final}(X) = k_{depth} \cdot f_{depth}(X) + k_{risk} \cdot f_{risk}(X) + k_{orientation} \cdot f_{orientation}(X) \quad (2)$$

2.5 Validation Method

We performed a retrospective validation of our method. For each case, we compared our solution with the position of the electrode that has been implanted in the patient's head, used as the reference trajectory. This actual trajectory was segmented from the post-operative CT images.

For each case, we computed the angle between our proposed optimal electrode trajectory, and the reference trajectory, in order to compare them. We also computed the scores of both trajectories for all of the individual soft constraints and the aggregative soft constraint, *i.e.* the result of their respective cost functions, in order to quantify their quality regarding the rules set with the neurosurgeons.

3 Results

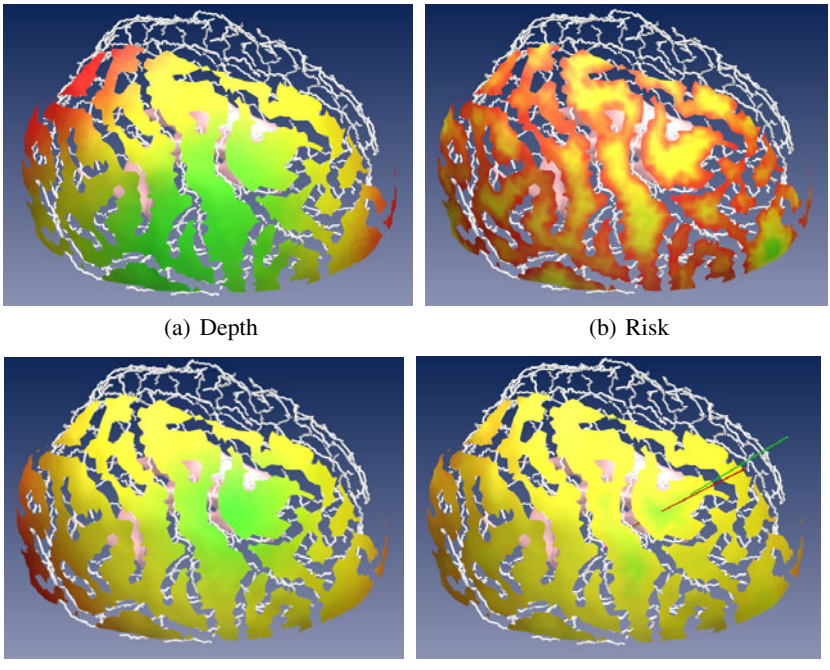
We performed a retrospective study on 8 cases, constituted by 4 patients who each had a bilateral STN implantation. We treated each side as a separate case for our study. For each case, we performed the registrations and segmentations described in Section 2.2 on preoperative and postoperative images, and we obtained the necessary structures and the targets. Then we launched the automatic planning application, using the constraints defined in Section 2.1, and weighting factors of 0.2 for k_{depth} (soft constraint #5),

Table 2. Comparison between trajectory produced by our planning and reference trajectory

Case #	1	2	3	4	5	6	7	8
a_{real} (degrees)	9.67	12.62	8.58	18.06	3.90	3.98	5.67	7.20
global score of T_{plan}	0.100	0.199	0.261	0.220	0.289	0.258	0.290	0.364
global score of T_{real}	0.295	0.391	0.431	0.429	0.417	0.467	0.473	0.504

0.4 for k_{risk} (#6), and 0.4 for k_{orient} (#7) to compute f_{final}. The weighting factors were defined by the neurosurgeons. However they obviously need to be refined more precisely, and this will be done in a future study.

We computed the angle a_{real} between trajectory T_{plan} produced by our path planning algorithm and the reference trajectory T_{real} of the electrode segmented from the post-operative CT. Results are shown on Table 2, along with the global scores (value of f_{final}) of each trajectory. As explained in Section 2.4, the global scores are real numbers between 0 and 1 (0 being the best score and 1 the worst), and represent in some way the percentage of satisfaction of the weighted soft constraints. We can notice that in all cases T_{plan} has a better global score than T_{real}.



(a) Depth (b) Risk

(c) Orientation (d) Aggregative: $k_{depth} = 2, k_{risk} = 4, k_{orient} = 4$

Fig. 2. Color maps of the soft constraints: best zones are in green and worst are in red (case #1). In snapshot (d) the red line is the computed trajectory T_{plan}, and the green line is the reference trajectory T_{real}.

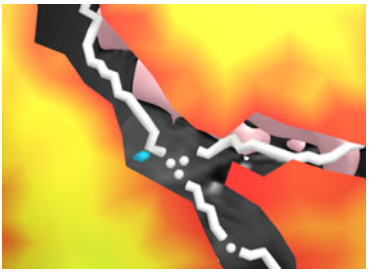


Fig. 3. Detail of the risk map. The target is in blue, the sulcal vessels in white, and the pink shape in the back is a part of the ventricles. The background is the grayscale MRI.

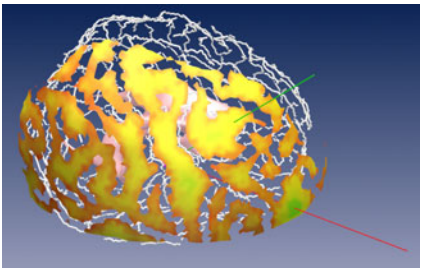


Fig. 4. Color map of the aggregative soft constraint, for different values of the weighting factors (case #1): $k_{depth} = 1, k_{risk} = 8, k_{orient} = 1$. T_{plan} is at another location.

Fig.2 shows snapshots of our software. It displays the possible insertion zones computed during the first step of our algorithm: elimination of the areas that would not satisfy the strict constraints. The points of the possible insertion zones are then colored, during the second step of our algorithm, according to their scores regarding each soft constraint (2(a),2(b),2(c)). The color map of the global aggregative soft constraint is shown on Fig.2(d). On this figure, the trajectory computed as optimal T_{plan} using the chosen weighting factors is shown as a red line, and the trajectory actually performed T_{real} as a green line. It can be noticed from this figure that very few areas are green, *i.e.* acceptable. Most of the areas are yellow or orange/red, *i.e.* medium or poor candidates according to the defined constraints and the chosen weighting factors.

Fig.3 shows a detail of the color map of the risk constraint. The map has a border where a trajectory towards the target (in blue) would meet any obstacle: here an approximation of a sulcal vessel represented by the white shape.

The experiments were performed on a 15" laptop, with a Dual Core CPU at 2.26 GHz and 4Go RAM, equipped with a NVIDIA GeForce Go 9300M GS GPU which is used to speed up occlusion queries. The constraint solving process and the generation of the colored maps and the optimal trajectory took less than 2 mn for each patient.

4 Discussion

We proposed a method able to provide valuable information for the neurosurgeon in the form of colored maps, that allow to see very quickly the individual scores of the possible insertion areas for each of the defined rules, facilitating the decision making. This method also computes an optimal path according to those rules, *i.e.* the path that has the best global score, result of the aggregative cost function, according to the chosen weightings. The computed angle a_{real} demonstrated that the trajectories performed by the expert neurosurgeon were not so far from our proposed path (average of 8.71 degrees), but not exactly the same. However in this kind of retrospective study it is difficult to compare the results with the terrain truth, as the trajectories that were actually used might not be the optimal ones. That is why we used constraints and weighting factors defined by the neurosurgeon, and computed the scores from them.

We can analyze the results in different ways. The weighting factors in the aggregative cost function might have not been chosen at best, and might need to be refined to obtain a better fitting with their trajectory. Indeed, in many cases (5 over 8) the scores of constraint #6 (risk) are better for T_{plan}, and the scores of constraint #7 (orientation) are better for T_{real}, suggesting that the optimization of the orientation of the electrode relatively to the axis of the target may be of greater importance for the neurosurgeons than they thought when setting the weighting factors. However we can also say on the contrary that, as the rules and the weighting factors we used were defined by the neurosurgeons, the trajectories we plan better fit their theoretical criteria (the global scores of T_{plan} are better than the scores of T_{real}), and maybe the planning tool they used in clinical routine did not provide them sufficient information and visibility for a correct selection.

An example of another possible computation with different weighting factors is shown on Fig 4 ($k_{depth} = 1, k_{risk} = 8, k_{orient} = 1$). On this figure, the green line representing T_{plan} is located in another place. On this example, we can notice that there is one main green zone (where the optimal trajectory is located). The greatest part of the areas is orange/red, because we gave a higher weight to the risk constraint that colors in orange/red the borders of the possible insertion zone. This example, where the result differs completely from the one we used in our validation, illustrates how the result is highly dependent on an accurate choice of the weighting factors. That is why we plan in our future works to refine them.

We might also discover in the future that a constraint is implicitly used by the neurosurgeon but has not been expressed so far. In that case, one great advantage of our approach is the modularity. The new constraint only has to be translated into a cost function, formalized, and simply added into the constraints XML file, and it will be automatically taken into account in the next planning.

5 Conclusion

We described an approach using a geometric constraint solver fed with two types of input data: the formalization of the rules governing DBS planning and patient-specific images, to automatically compute an optimal placement for an electrode in the framework of assistance to DBS planning. The results we obtained show that the solutions proposed by our solver differed little from the solutions that were actually performed in clinical routine, but they had better scores regarding the rules defined by the surgeon themselves.

In the future we plan to feed the solver with other types of information, such as clinical scores, or predictions of the deformation of the electrode. Further clinical evaluation with more surgical cases will also be performed, including other types of targets: globus pallidus internus (GPi) and ventral intermediate nucleus of the thalamus (VIM). Further works would also include the addition of new constraints that could be discovered and enunciated, especially when new targets will be considered. The modularity of our xml rule-based system will make that addition quite fast and convenient, as long as the new rules are clearly identified.

Acknowledgments

We would like to thank Dr. Eduardo Pasqualini from the Department of Neurosurgery at Pontchaillou University Hospital Rennes, Florent Lalys and Alexandre Abadie from VisAGeS/U 746 - University of Rennes 1, and Markus Engel and all MITK-team from DKFZ Heidelberg, for their useful advices and help.

References

1. D'Haese, P.F., Cetinkaya, E., Konrad, P.E., Kao, C., Dawant, B.M.: Computer-aided placement of deep brain stimulators: from planningto intraoperative guidance. IEEE Trans. Med. Imaging 24(11), 1469–1478 (2005)
2. Altrogge, I., Kröger, T., Preusser, T., Büskens, C., Pereira, P., Schmidt, D., Weihusen, A., Peitgen, H.: Towards optimization of probe placement for radio-frequency ablation. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 486–493. Springer, Heidelberg (2006)
3. Lung, D., Stahovich, T., Rabin, Y.: Computerized planning for multiprobe cryosurgery using a force-field analogy. Comput. Meth. Biomech. Biomed. Eng. 7(2), 101–110 (2004)
4. Adhami, L., Coste-Manière, E.: Optimal planning for minimally invasive surgical robots. IEEE Transactions on Robotics and Automation 19(5), 854–863 (2003)
5. Brunenberg, E., Vilanova, A., Visser-Vandewalle, V., Temel, Y., Ackermans, L., Platel, B., ter Haar Romeny, B.: Automatic trajectory planning for deep brain stimulation: A feasibility study. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 584–592. Springer, Heidelberg (2007)
6. Baegert, C., Villard, C., Schreck, P., Soler, L.: Multi-criteria trajectory planning for hepatic radiofrequency ablation. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 584–592. Springer, Heidelberg (2007)
7. Baegert, C., Villard, C., Schreck, P., Soler, L.: Trajectory optimization for the planning of percutaneous radiofrequeny ablation on hepatic tumors. Computer Aided Surgery (2007)
8. Lalys, F., Haegelen, C., Abadie, A., Jannin, P.: Post-operative assessment in deep brain stimulation based on multimodal images: registration workflow and validation. In: Medical Imaging: Vis., Image-Guided Procedures, and Modeling, vol. 7261(1), 72612M. SPIE, San Jose (2009)

FEM Based 3D Tumor Growth Prediction for Kidney Tumor

Xinjian Chen, Ronald Summers, and Jianhua Yao

Radiology and Imaging Sciences Department, Clinical Center, National Institute of Health, Bethesda, MD 20814, USA
Ph.: (301)594-8139; Fax: (301)480-9827
chenx6@mail.nih.gov

Abstract. It is important to predict the tumor growth so that appropriate treatment can be planned especially in the early stage. In this paper, we propose a finite element method (FEM) based 3D tumor growth prediction system using longitudinal kidney tumor images. To the best of our knowledge, this is the first kidney tumor growth prediction system. The kidney tissues are classified into three types: renal cortex, renal medulla and renal pelvis. The reaction-diffusion model is applied as the tumor growth model. Different diffusion properties are considered in the model: the diffusion for renal medulla is considered as anisotropic, while those of renal cortex and renal pelvis are considered as isotropic. The FEM is employed to simulate the diffusion model. Automated estimation of the model parameters is performed via optimization of an objective function reflecting overlap accuracy, which is optimized in parallel via HOPSPACK (hybrid optimization parallel search). An exponential curve fitting based on the non-linear least squares method is used for multi-time point model parameters prediction. The proposed method was tested on the seven time points longitudinal kidney tumor CT studies from two patients with five tumors. The experimental results showed the feasibility and efficacy of the proposed method.

Keywords: Tumor Growth Prediction, Finite Element Method, Segmentation, Kidney Tumor.

1 Introduction

Kidney cancer is among the 10 most common cancers in both men and women. Overall, the lifetime risk for developing kidney cancer is about 1 in 75 (1.34%) [1]. It is important to predict the kidney tumor growth rate in clinical research so that appropriate treatment can be planned.

During the last three decades, the methods for simulating tumor growth have been extensively studied. The representative methods include mathematical models [2, 3], cellular automata [4], finite element [3, 5] and angiogenesis based methods [6]. However, most of these methods are focused on brain tumor growth prediction. Only a few can be found for organs in the body region. Pathmanathan et al. [7] proposed to use the finite-element method and nonlinear elasticity to build a 3-D patient specific breast model, which was used to predict the tumor location. However, this method was not for tumor growth prediction.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 159–168, 2010.

In this paper, we propose a tumor growth prediction system for kidney tumor based on finite element method (FEM). The kidney tissues are classified into three main types: renal cortex, renal medulla and renal pelvis (or collecting system). Based on Chandarana et al. [8], different diffusion properties are considered for the kidney: the renal cortex and renal pelvis are considered to be isotropic while renal medulla be anisotropic. The reaction-diffusion model is applied here to model the tumor growth and the FEM is applied to simulate this diffusion process.

The growth rate of renal tumors can be very slow. Kassouf et al. [9] followed up 24 patients over a period of average 24 months, and found noticeable tumor growth in only five patients during the surveillance period. Since the growth is slow, longitudinal studies over a long period of time are required to monitor the disease progress. A tumor growth prediction based on longitudinal studies over a short time period can help the physicians to plan the treatment in the early stage.

Our main contributions are summarized as follows:

- To the best of our knowledge, this is the first kidney tumor growth prediction system. Three different tissue types and diffusion properties are considered in the model. Reaction-diffusion model is applied to model the tumor growth and finite element method is employed to simulate the diffusion.
- Automated estimation of the model parameters is performed via optimization of an objective function reflecting overlap accuracy, which is optimized in parallel via HOPSPACK (hybrid optimization parallel search).
- An exponential curve fitting based on the non-linear least squares method is used to predict model parameters in longitudinal studies.

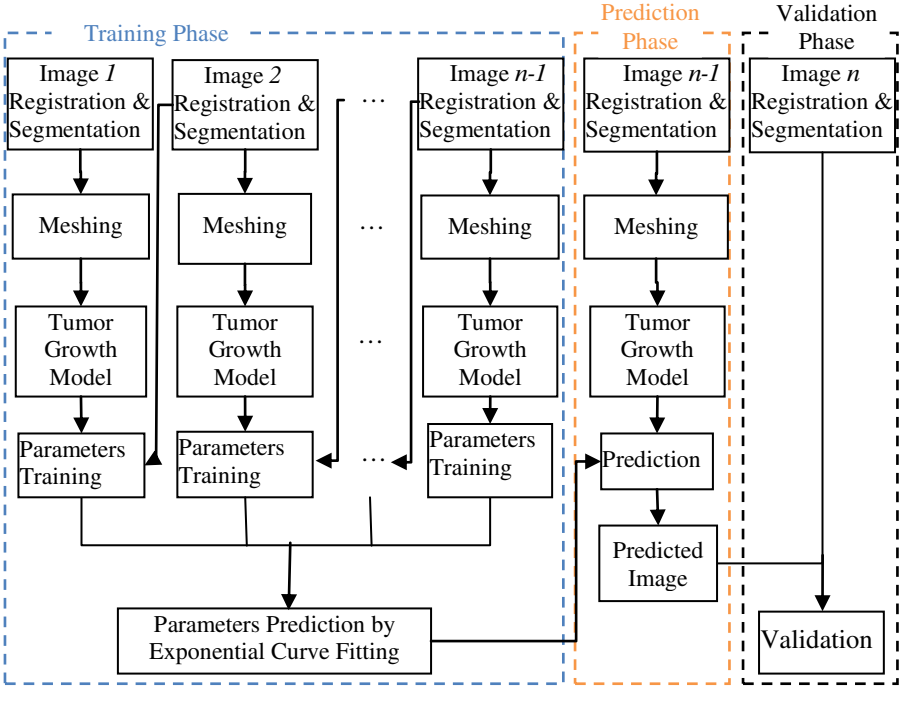


Fig. 1. The flowchart of tumor growth prediction system

2 FEM Based Tumor Growth Prediction

The proposed tumor growth prediction system consists of three main phases: training, prediction and validation. The flowchart is shown in Fig 1. Suppose the longitudinal study has n time points. For the purpose of validation, we use first $n-1$ time point images for training, predict the tumor status at the n^{th} time point, and validate with the n^{th} images. In clinical practice, all n images are used to train the model parameters and predict the tumor status at a future time point. The training phase is composed of five steps. First, image registration and segmentation are applied to the kidney images. Second, tetrahedral meshes are constructed for the segmented kidney and tumors, respectively. Third, the reaction-diffusion model is applied as the tumor growth model, and FEM is used to solve this partial differential equation (PDE). Fourth, the parameters of the tumor growth model are estimated using the two studies at different time points (preferably two adjacent time points). Fifth, after computing the parameters based on the first $n-1$ image, the model parameters for prediction at the n^{th} time point are estimated by an exponential curve fitting based on the non-linear least squares method. In the prediction phase, the estimated growth parameters are applied to the tumor growth model based on image $n-1$ to compute the predicted result for time point n. In the validation phase, the prediction result is validated by comparing with image n.

2.1 Registration and Image Segmentation

The baseline study is used as the reference study and all other studies are registered to it via a rigid transformation. Then the kidney is segmented by the graph-cut oriented active appearance method [10]. This method synergistically combines the active appearance model, live-wire and graph cut methods to take advantage of their complementary strengths. The details can be seen in [10]. After the kidney is segmented, the tumors, renal cortex and renal pelvis are manually segmented and the remaining tissues are treated as renal medulla.

2.2 Meshing

A tetrahedral mesh is built for the segmented tissues. The full meshing procedure is composed of the following three steps:

1) A surface mesh is first generated for the segmented tissue (kidney and tumors) by the marching cube algorithm [11].
2) This surface mesh is then decimated by the ISO2Mesh method [12].
3) The volumetric mesh is finally generated from the surface mesh also by the ISO2Mesh method [12].

2.3 Tumor Growth Model

The reaction-diffusion model is adopted to model the growth and spreading of tumor cells in the kidney. The reaction-diffusion model [3] was first proposed in chemistry, and widely used in biology, geology, physics and ecology. The model is defined as follows:

$$\frac{\partial c}{\partial t} = -div(-D\nabla c) + S(c,t) - T(c,t) \tag{1}$$

where c represents the tumor cell density, D is the diffusion coefficient of tumor cells, $S(c,t)$ represents the source factor function which describes the proliferation of tumor cells and $T(c,t)$ is used to model the efficacy of the tumor treatment.

Since our purpose is only to predict the tumor growth before treatment, the treatment term $T(c,t)$ is omitted. The source factor $S(c,t)$ can be modeled using Gompertz law [3], which is defined as follows,

$$S(c,t) = \rho\, c \ln(\frac{C_{\max}}{c}) \tag{2}$$

where ρ is the proliferation rate of tumor cells, C_{max} is the maximum tumor cell carrying capacity of the kidney tissue. According to [3], C_{max} is set to 3.5×10^4 Cells mm^{-3}.

Combining Equ. (2) and (1) and omitting *T(c,t)*, we can get,

$$\frac{\partial c}{\partial t} = -div(-D\nabla c) + \rho c \ln(\frac{C_{\max}}{c}) \tag{3}$$

Clatz et al. [3] assumed the anisotropic ratio of diffusion is the same for water molecules and tumor cells so that, and computed the brain tumor diffusion coefficients based on the diffusion tensor image. However, further experiments are needed to validate the hypothesis. Based on Chandarana et al. [8] and Notohamiprodjo et al. [14], in this paper the diffusion in the renal cortex and renal pelvis are considered to be isotropic, while that in renal medulla be anisotropic and the diffusion in the radial direction is faster than other directions. The diffusion properties of different tissues are listed in Table 1. It is important to note that in the diffusivity matrix D_m of medulla, the diffusivity in the radial direction is set as λ times of those in other directions.

Table 1. Diffusion properties (D) of different kidney tissues

Tissue	Diffusivity (10^{-3} mm^2s^{-1})
Renal cortex	D_c (isotropic)
Renal medulla	D_m (anisotropic)
Renal pelvis	D*p* (isotropic)

The finite element method (FEM) is used to solve the PDE in the above reaction-diffusion model. Based on the Galerkin method [15], the continuous problem can be converted to a discrete problem in a subvectorial space of finite dimension. In principle, it is the equivalent of applying the method of variation to a function space, by converting the equation to a weak formulation [15]. The details of implementation of reaction-diffusion model by FEM can be found in [15].

2.4 Tumor Growth Model Parameters Training

In our tumor growth model, D_c, D_m, D_p, ρ are the parameters need to be estimated. The optimal set of tumor growth parameters for a particular patient is not known; it must be estimated from the patient's image. The optimizing of the tumor model parameters is based on the hypothesis that the optimal tumor parameters minimize the discrepancies between the simulated tumor image and patient tumor image. We achieve this goal by solving the following optimization problem,

$$\theta^* = \arg\min_{\theta} E(\theta) \tag{4}$$

Where $\theta = \{D_c, D_m, D_p, \rho\}$, E is the objective function. Many criteria can be used for constructing function E, such as the overlap accuracy, feature based similarity and smoothness of the registration in [18, 19]. Here, we only use the overlap accuracy. As mentioned earlier, in this paper the first *n-1* studies are used for model parameters training. The parameters are trained in pair using consecutive study i and $i+1$. Our objective function is defined as follows,

$$E(\theta) = \sum_{i=1}^{n-2} w \cdot (1 - TPVF(I_{i,\theta}, I_{i+1})) + (1-w) \cdot FPVF(I_{i,\theta}, I_{i+1}) \tag{5}$$

Where, I_{i+1} is used here as the validation tumor image, $I_{i,\theta}$ is the predicted tumor image using θ based on image I_i. For giving parameters θ, we can compute the density for each node in the tetrahedron element based on the tumor growth model. A linear interpolation is then used to obtain the density for every point inside the tetrahedron. The threshold method is applied to detect the tumor. Tracqui *et al.* [17] suggested an 8000 cell mm^{-3} threshold of detection for an enhanced CT scan. This value is also applied in this paper. w is the weight for *TPVF* (in this paper w=0.5). TPVF (true positive volume fraction) indicates the fraction of the total amount of tissue in the true delineation; FPVF (false positive volume fraction) denotes the amount of tissue falsely identified, which are defined as follows,

$$\text{TPVF} = \frac{C_{TP}}{C_{td}} \tag{6}$$

$$\text{FPVF} = \frac{C_{FP}}{U_d} \tag{7}$$

Where, U_d is assumed to be a binary scene with all voxels in the scene domain C set to have a value 1, as shown in Fig. 2, more details can seen in [13].

The optimization of Eq. (4) is not a trivial task due to the defined objective function contains discontinuities. However, Pattern Search methods are suitable for such problems [22, 23]. They are directional methods that make use of a finite number of directions with appropriate descent properties. We apply a hybrid optimization parallel search method, called HOPSPACK [22], which takes advantage of parallel platforms. HOPSPACK comes with an asynchronous pattern search solver that handles general

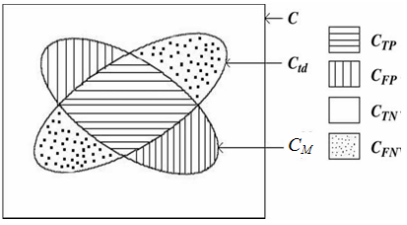


Fig. 2. Illustration of the accuracy factors for delineation for a binary case. Here, C_{td} is orresponding scene of 'true' delineation, C_M is the delineation result by method *M*.

optimization problems with linear and nonlinear constraints, and continuous and integer-valued variables. Due to the complicated form of our objective function, it is not guaranteed that a global optimum exists. We set an iterations limit of 200 in the optimization process. Usually a minimum was reached and the optimization terminated before reaching this limit.

We assumed here the diffusion properties for the tissues are not changed over the time, while the proliferation rate ρ could be changed. Suppose $\rho_1, \rho_2, \ldots, \rho_m$ are the estimated tumor growth parameters, which corresponding to time point $t_1, t_2, \ldots, t_m$. Then we need to predict the proliferation rate ρ for the future time point. West et al [20] shows that, regardless of the different masses and development times, mammals, birds, fish and mollusks, all share a common tumor growth pattern. Provided that masses and growth times for the different organisms are properly rescaled, the same universal exponential curve fits their ontogenetic growth data. In this paper, we also assume the tumor growth follows this exponential law, which is defined as follows:

$$\rho = a * \exp(b * t) + c * \exp(d * t) \tag{8}$$

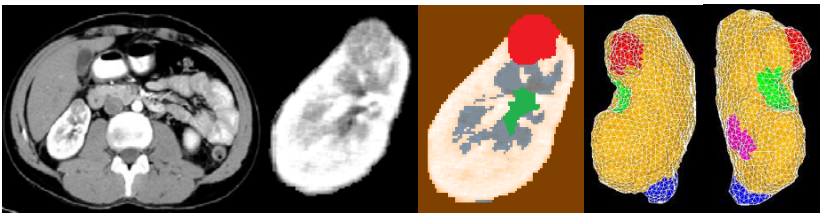
Where, *a, b, c, d* are the growth coefficients. The non-linear least squares method is used for curve fitting. The detail can be seen in [24].

3 Experimental Results

We tested the proposed methods on two longitudinal studies of kidney tumors. The contrast enhanced CT images in arterial phase were used. Both studies had 7 time points images scanned at regular intervals over 3 to 4 years. Three kidney tumors were monitored for study #1, and two kidney tumors were monitored for study #2. The CT images were acquired from GE LightSpeed QX scanner with the slice spacing = 5.00 mm and pixel size = 0.78x0.78 mm^2. All images were segmented manually by an expert to generate the ground truth.

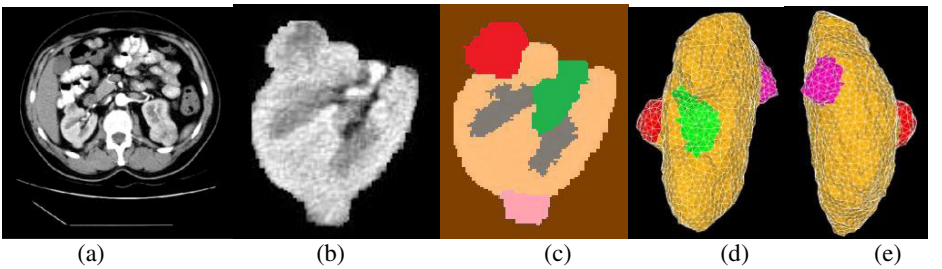
Figs. 3 and 4 show the segmentation results and meshes for both patients, respectively. A mesh consisting of 7217 nodes and 40996 tetrahedra was generated for the first study, and 6666 nodes and 37857 tetrahedra for the second study.

As both studies had 7 time point images, the training of tumor growth model parameters were done on their first six images. As mentioned earlier, we assumed the diffusion properties for the tissues are not changed over the time, while the proliferation rate ρ could be changed. The diffusivities (D_c, D_p) are isotropic, and the trained



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

Fig. 3. Segmentation and meshing results for the first study. (a) Original image, (b) the segmented kidney, (c) the segmented tissues and one tumor (cortex: orange, medulla: black, pelvis: green, tumor: red), (d) and (e) two different views of volumetric mesh (cortex: orange, pelvis: green, tumors: red, pink and blue, medulla is invisible because it is inside).



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

Fig. 4. Segmentaion and meshing results for the second study. (a) Original image, (b) the segmented kidney, (c) the segmented tissues and two tumors (cortex: orange, medulla: black, pelvis: green, tumor: red and pink), (d) and (e) two different views of volumetric mesh (cortex: orange, pelvis: green, tumors: red and pink, medulla is invisible because it is inside).

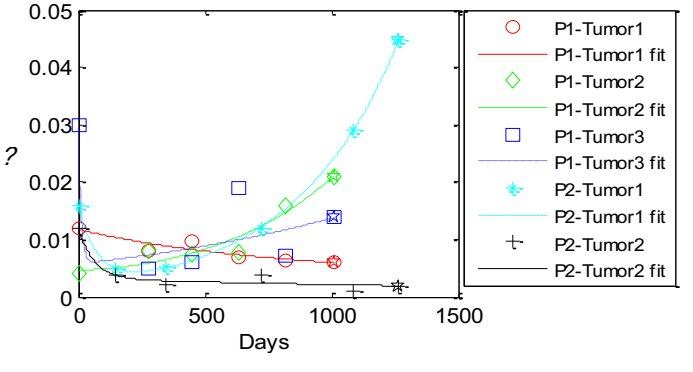


Fig. 5. Parameter ρ curve fit by eq. (8) for all 5 tumors. The fit was based on the 5 estimated values, the 6th value (overlap with ☆) was used for prediction also shown on the figure.

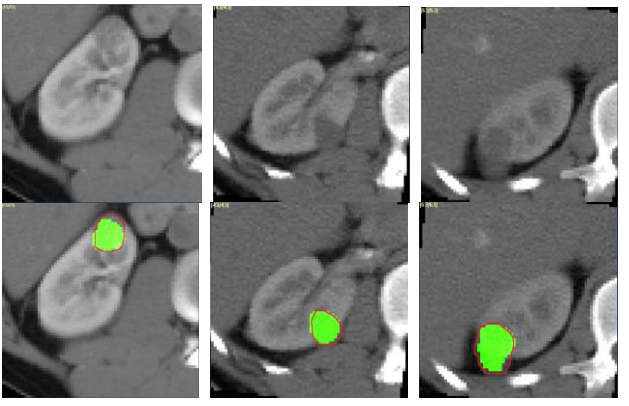
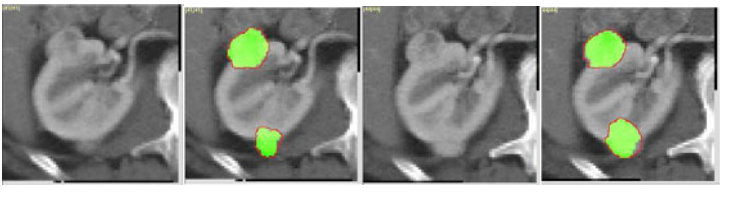


Fig. 6. Results of the tumor growth prediction on three slices for the first study. Top row shows the original images, bottom row shows the prediction results by green overlaid on the original images. Red line represents the manually segmented tumor results. Different column shows different tumor.



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

Fig. 7. Results of the tumor growth prediction on two slices for the second study. (a) and (c) are the original images. (b) and (d) shows the prediction results by green overlaid on (a) and (c), respectively. Red line represents the manually segmented tumor results.

values are ($3.2\text{x}10^{-5}$, $3.8\text{x}10^{-5}$) and ($3.2\text{x}10^{-5}$, $3.8\text{x}10^{-5}$) ($\text{mm}^{-2}\text{s}^{-1}$) for study#1 and study #2, respectively. D_m is anisotropic, the trained values in radial and other directions are $3.5\ \text{x}10^{-5}$, $2.9\ \text{x}10^{-5}$ $\text{mm}^{-2}\text{s}^{-1}$ (λ=1.207) for study #1; and $3.15\ \text{x}10^{-5}$, $2.6\ \text{x}10^{-5}$ $\text{mm}^{-2}\text{s}^{-1}$ (λ=1.212) for study #2, respectively. These values are consistent with the diffusion value in [8, 14] (within one order of magnitude). The trained ρ for five tumors (3 for patient #1 and 2 for patient 2) are shown in Fig. 5. The curve fitting based on Eq. (8) using non-linear least squares method was applied to these data. The computed ρ based on curve fit was applied on the 6^{th} image to predict the tumor and validated with the 7^{th} image. Fig. 6 and 7 show the predicted results for these two studies. We can find the predicted results are quite good.

As for quantitative evaluation, the TPVF and FPVF [13] are used to show the accuracy of the proposed method, the results are also shown in Table 2. The average TPVF and FPVF on all tumors is quite good, 91.4% and 4.0%, respectively. We also

Table 2. Volume difference, TPVF and FPVF for each tumor in the two studies

	Study #1			Study #2	
	Tumor1	Tumor2	Tumor3	Tumor1	Tumor2
Volume difference	4.1%	5.1%	4.8%	4.5%	5.7%
TPVF	91.5%	90.9%	91.6%	92.1%	90.8%
FPVF	4.6%	4.2%	3.7%	3.5%	3.8%

list the volume difference between the predicted results and the manually segmented results in Table 2. The average volume difference is about 4.8%.

4 Conclusions and Discussions

Tumor growth prediction provides useful information in tumor treatment planning. Based on the prediction, the physician can choose among the treatments such as radiotherapy, chemotherapy or surgery. In this paper, a FEM based 3D tumor growth prediction using longitudinal studies of kidney tumor was proposed. The proposed method was tested on two longitudinal studies with seven time points on five tumors. The preliminary experimental results proved the feasibility and efficacy of the proposed system. The proposed system can be applied to predict tumor growth in other organs by modifying the tissue and diffusion properties accordingly.

In this paper, we adopted a widely used reaction-diffusion model as the tumor growth model. A more complex model can be considered, such as coupling diffusion with biomechanical model [3, 19]. This will be investigated in the future work.

Acknowledgements. We would like to thanks BS. Francine Thomas, and MD. Suvimol Hill for their kindly help on the kidney anatomy.

References

1. American Cancer Society, http://www.cancer.org/docroot/cri/content/cri_2_4_1x_what_are_the_key_statistics_for_kidney_cancer_22.asp
2. Swanson, K., Bridge, C., Murray, J.D., Alvord, E.C.: Virtual and real brain tumors: Using mathematical modeling to quantify glioma growth and invasion. J. Neurol. Sci. 216(1), 1–10 (2003)
3. Clatz, O., Sermesant, M., Bondiau, P.Y., Delingette, H., Warfield, S.K., Malandain, G., Ayache, N.: Realistic simulation of the 3-D growth of brain tumors in MR images coupling diffusion with biomechanical deformation. IEEE Trans. Med. Imaging 24(10), 1334–1346 (2005)
4. Mallet, D.G., Pillis, L.G.D.: A cellular automata model of tumor-immune interactions. Journal of Theoretical Biology 239(3), 334–350 (2006)

5. Mohamed, A., Davatzikos, C.: Finite element modeling of brain tumor mass-effect from 3D medical images. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 400–408. Springer, Heidelberg (2005)
6. Lloyd, B.A., Szczerba, D., Székely, G.: A Coupled Finite Element Model of Tumor Growth and Vascularization. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 874–881. Springer, Heidelberg (2007)
7. Pathmanathan, P., Gavaghan, D.J., Whiteley, J.P., Chapman, S.J., Brady, J.M.: Predicting tumor location by modeling the deformation of the breast. IEEE Trans. Biomed. Eng. 55(10), 2471–2480 (2008)
8. Chandarana, H., Hecht, E., Taouli, B., Sigmund, E.E.: Diffusion Tensor Imaging of In Vivo Human Kidney at 3T: Robust anisotropy measurement in the medulla. Proc. Intl. Soc. Mag. Reson. Med. 16, 494 (2008)
9. Kassouf, W., Aprikian, A.G., Laplante, M., Tanguay, S.: Natural history of renal masses followed expectantly. J. Urol. 171, 111–113 (2004)
10. Anonymous
11. Lorensen, W., Cline, H.: Marching cubes: A high resolution 3-D surface construction algorithm. In: Proc. Siggraph 1987, C. Graphics, vol. 21, pp. 163–170 (July 1987)
12. Fang, Q.: ISO2Mesh: a 3D surface and volumetric mesh generator for matlab/octave, `http://iso2mesh.sourceforge.net/cgi-bin/index.cgi?Home`
13. Udupa, J.K., Leblanc, V.R., Zhuge, et al.: A framework for evaluating image segmentation algorithms. Computerized Medical Imaging and Graphics 30(2), 75–87 (2006)
14. Notohamiprodjo, M., Glaser, C., Herrmann, K.A., et al.: Diffusion Tensor Imaging of the Kidney With Parallel Imaging: Initial Clinical Experience. Investigative Radiology 43(10), 677–685 (2008)
15. Hanhart, A.l., Gobbert, M.K., Izu, L.T.: A memory-efficient finite element method for systems of reaction-diffusion equations with non-smooth forcing. Journal of Computational and Applied Mathematics 169, 431–458 (2004)
16. Laird, A.K.: Dynamics of tumor growth. Br. J. of Cancer 18, 490–502 (1964)
17. Tracqui, P., Cruywagen, G., Woodward, D., Bartoo, G., Murray, J., Alvord Jr., E.: A mathematical model of glioma growth: The effect of chemotherapy on spatio-temporal growth. Cell Proliferation 28(1), 17–31 (1995)
18. Zacharaki, E.I., Shen, D., Lee, S.-K., Davatzikos, C.: ORBIT: A Multiresolution Framework for Deformable Registration of Brain Tumor Images. IEEE Transactions on Medical Imaging 27(8) (August 2008)
19. Hoge, C., Davatzikos, C., Biros, G.: An image-driven parameter estimation problem for a reaction-diffusion glioma growth model with mass effects. J. Math. Biol. 56, 793–825 (2008)
20. West, G.B., Brown, J.H., Enquist, B.J.: A general model for ontogenetic growth. Nature 413, 628–631 (2001)
21. Guiot, C., Degiorgis, P.G., Delsanto, P.P., Gabriele, P., Deisboeck, T.S.: Does tumor growth follow a "universal law"? J. Theor. Biol. 225(2), 147–151 (2003)
22. Plantenga, T.D.: HOPSPACK 2.0 User Manual, TECHREPORT SAND2009-6265, Sandia National Laboratories, Albuquerque, NM and Livermore, CA (October 2009)
23. Gray, G.A., Kolda, T.G.: Algorithm 856: APPSPACK 4.0: Asynchronous parallel pattern search for derivative-free optimization. ACM Transactions on Mathematical Software 32, 485–507 (2006)
24. Kelley, C.T.: Iterative Methods for Optimization. SIAM Frontiers in Applied Mathematics 18 (1999), ISBN: 0-89871-433-8

Adaptive GPU Ray Casting Based on Spectral Analysis

Stefan Suwelack, Eric Heitz, Roland Unterhinninghofen, and Rüdiger Dillmann

Institute of Anthropomatics (IFA) Humanoids and Intelligence Systems Laboratories (HIS), Karlsruhe Institute of Technology (KIT), Germany
suwelack@kit.edu

Abstract. GPU based ray casting has become a valuable tool for the visualization of medical image data. While the method produces high-quality images, its main drawback is the high computational load. We present a novel adaptive approach to speed up the rendering. In contrast to well established heuristic methods, we use the spectral decomposition of the transfer function and the dataset to derive a suitable sampling criterion. It is shown how this criterion can be efficiently incorporated into an adaptive ray casting algorithm. Two medical datasets, which each represent a typical, but different material distribution, are rendered using the proposed method. An analysis of the number of sample points per ray reveals that the new algorithm requires 50% to 80% less points compared to a non-adaptive method without any quality loss. We also show that the rendering speed of the GPU implementation is greatly increased with reference to the non-adaptive algorithm.

Keywords: volume rendering, gpu raycasting, adaptive sampling, spectral analysis.

1 Introduction

The 3D visualization of computer tomographic (CT) data by means of real-time volume rendering techniques is widely used in clinical diagnostics and medical education [11]. A very powerful and popular approach to do so is to perform ray casting on the graphics processing unit (GPU) [7]. This method allows the interactive rendering of moderately sized datasets on a standard desktop PC.

Many different approaches have been proposed to increase the efficiency of GPU based ray casting. View-dependent octree data structure have been proposed to reduce the memory consumption on the GPU side [4] [2]. The computational load can be reduced by adapting either the number of rays (image based adaptivity) or the number of sample points for each ray (object based adaptivity). While image based adaptive methods can be effective in combination with level-of-detail methods [8] we will focus here on object based adaptive sampling.

Even very early GPU based ray casting algorithms employed empty space skipping and early ray termination to reduce the sample points per ray [7]. A full adaptive sampling algorithm based on a so-called importance volume was

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 169–178, 2010.

proposed by Roettger et al. [10]. Recently presented ray casting engines often rely on multi-resolution data structures and therefore naturally embed full adaptive sampling. However, very little attention has been paid to the criteria that control the sampling distance. The sampling distance is either chosen according to a given multi-resolution representation of the volume [8] or heuristic properties such as the gradient of the dataset [6].

Mathematically sound adaptivity criteria can be obtained by applying a Fourier analysis [1]. The sampling distance then follows through the Nyquist-Shannon sampling theorem. If non-linear transfer functions are used (which is the case in medical applications), it is not sufficient to apply the spectral decomposition just to the scalar data $f(x)$ [5]. In order to capture high-frequency characteristics of the transfer function $g(x)$ the complete signal $(g \circ f)(x) = g(f(x))$ has to be analyzed instead. However, the transfer function usually changes during the rendering to highlight different anatomical regions of interest. Therefore, an online Fourier decomposition of the whole dataset would be necessary in order to determine the Nyquist rate as soon as the transfer function changes. This is not feasible due to the large computational effort associated with this analysis.

In order to alleviate this constraint, several authors proposed a technique called pre-integration of the transfer function [3] [12]. Based on the assumption that the scalar function is linear between the two sample points s_1 and s_2, the volume integral is determined analytically in this interval. The integral values are pre-computed and can be accessed via a lookup table for all values of s_1 and s_2. In this way high frequency behavior of the transfer function can be absorbed and the sampling rate can be reduced [5].

Unfortunately, the approach does not reduce the sampling rate where small changes of g actually reduce the maximal frequencies of the complete signal $g \circ f$. Furthermore, the signal reconstruction with the ray casting technique introduces further errors that deteriorate the quality of the pre-integrated solution. This issue will be discussed in detail in the next section. Bergner et al. showed a way to overcome these limitations [1]. Drawing from well established results in the field of signal processing they derive a mathematically rigorous estimation for the spectral decomposition of $g \circ f$. The key result establishes an estimate for the Fourier analysis of $g \circ f$ based on the Fourier analysis of g and the local gradient $f'(x)$.

We build upon the work of Bergner et al. and present a GPU based ray casting algorithm that incorporates object-space adaptive sampling based on a spectral analysis of $g \circ f$. In Section 2 we describe the errors that are introduced when applying ray casting for 3D image reconstruction. We subsequently explain the relation between the Nyquist rate and the sampling frequency. In Section 3 we show how different local behavior of the transfer function can be incorporated in the approach. Most importantly, we present how dynamic transfer functions can be used in the adaptive framework with minimal computational effort. A memory efficient implementation of the method is presented in Section 4. We also analyze the 3D rendering of typical medical CT datasets. The performance of the new method is compared to the theoretical optimal adaptive algorithm and

to a non-adaptive approach. We conclude with a summary of key contributions of this paper and give a short outlook for future work in Section 5.

2 Application of Sampling Theory to Ray Casting

A CT dataset can be regarded as a set of samples that have been taken from a band limited, continuous signal. According to the Nyquist-Shannon sampling theorem, the sampling frequency must be at least twice as high as the highest frequency of the original signal. Thus, for an isotropic sample distance of T between the voxels, the maximal frequency v_f that is represented in the dataset is given by $v_f = 1/2T$. The signal can be perfectly reconstructed by sampling the dataset once per voxel and then performing a convolution with a sinc filter kernel. However, additional errors are introduced, if ray casting is used to reconstruct the signal. These errors are discussed in the following paragraph by means of 1D examples. However, all results can be easily generalized to the 3D case.

The ray casting method essentially solves the rendering integral by adding the volume densities along each viewing ray. Usually tri-linear interpolation is employed to calculate the density values at the sample points. From a mathematical perspective, this approach corresponds to a reconstruction with a linear ('tent') filter kernel [9]. The errors that occur due to the reconstruction with the simplified filter kernel are referred to as filtering artifacts . We note two important properties of this error source. First of all, the filtering artifacts cannot be reduced by raising the sampling rate. This is quickly illustrated for a 1D signal (see Fig. 1). If the samples are always taken exactly in the center of each voxel, there is no additional benefit from sampling the signal with a higher frequency than once per voxel. Secondly, in contrast to the sinc filter, which has an infinite support, the tent filter's support is limited to adjoining voxels. Thus, the required sampling rate only depends on the local behavior of the material density.

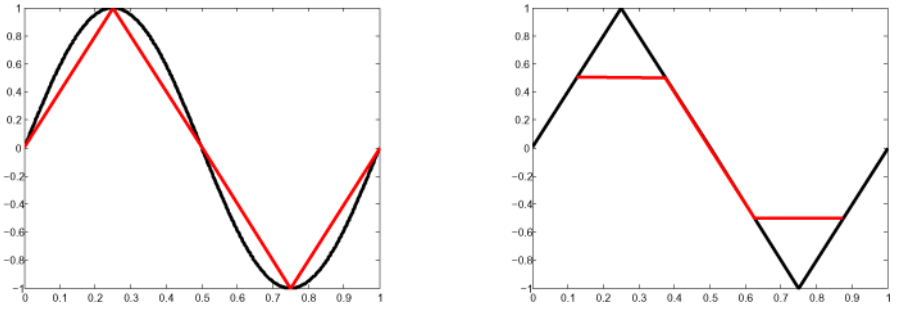


Fig. 1. A sinusoid signal (black) is reconstructed with a ray casting method (red). The filter kernel creates filtering artefacts (left). Samples taken with an offset introduce another error source (right), which can be reduced by performing oversampling.

In the previous discussion we assumed that the sample points are always placed in the centre of each voxel. However, during ray casting the samples are taken along the viewing ray at different positions that not necessarily coincide with the midpoint of the voxels. Fig. 1 illustrates the error that emanates from this procedure. The size of the error depends on the frequency of the signal, the sampling rate and the actual sample point, which cannot be known a-priori. Although the error scales with the Nyquist frequency of the signal, the simple 1D example (Fig. 1) already suggests that it is often not sufficient to sample with the Nyquist rate. This is in fact the case for volume rendering of medical datasets. Previously published results suggest that an oversampling by a factor of two to four with respect to the Nyquist frequency is enough to remedy the problem [10].

In order to generate visually pleasant and meaningful images, a transfer function g is used to assign optical properties to the Hounsfield values stored in the CT dataset. Thus, the sampling frequency does not only depend on the data values f, but rather on the composite function $g \circ f$. This is especially important, if the transfer function has high frequency components.

3 Spectral Analysis

The direct spectral decomposition of the composite function $g \circ f$ (e.g. by the discrete Fourier transform) is a very computational intensive procedure. If the transfer function does not change during the rendering this analysis can be done in a pre-processing step. However, in medical applications the transfer function is usually changed during the rendering to browse through different anatomical structures and regions of interest. In this section we present an efficient approximation for the spectral decomposition that can be evaluated with very little computational effort. This allows the use of frequency based adaptivity criteria even for transfer function that change dynamically during the rendering.

3.1 Spectral Analysis of the Composite Function

Bergner et al. showed that the maximal frequency v_h of the composite function $g \circ f$ can be approximated by

$$v_h = v_g \max_x |f'(x)|\,, \tag{1}$$

because the energy of the signal decays exponentially for frequencies above this estimation [1]. Here, v_g denotes the maximum frequency of g. In accordance with these results our experiments showed that no visible artifacts occur even when no additional oversampling is used.

While the estimate above holds for all volume rendering techniques, it can be significantly enhanced for the ray casting technique due to the locality of the reconstruction filter. When approximating the frequency to determine the step size from sample point s_1 to sample point s_2 it is obvious that the maximal

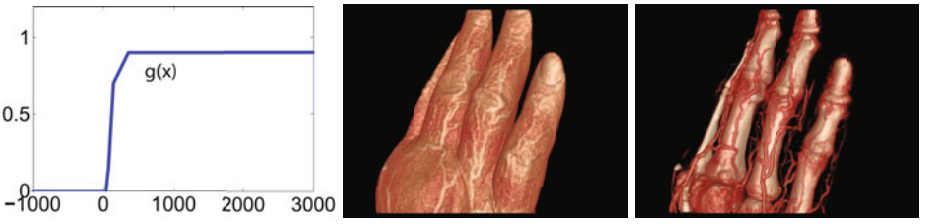


Fig. 2. A typical non-stationary opacity transfer function used for medical visualization (left). The transfer function is shifted along the x-axis to browse through different anatomical structures of a human hand (middle, right).

gradient of $f'(x)$ should be determined over the interval $[s_1, s_2]$ and not over the whole volume. Additionally, we propose to determine the maximal frequency of g not over the entire range of Hounsfield values, but rather over these values $I = [s_{\min}, s_{\max}]$ that actually occur in $[s_1, s_2]$. The benefits of this approach become quickly apparent when looking at the transfer functions that are typically used for medical visualization (see Fig. 2). There are large areas where the transfer function has minimal variations whereas it changes substantially around a certain threshold value. The step size can therefore be reduced if all values of f in $[s_1, s_2]$ only induce little changes in the transfer function. Using the approximation given by equation (1) the sampling period is given by

$$T = \frac{1}{2v_g(I) \cdot \max |f'(x)|} \tag{2}$$

where $v_g(I)$ is the maximal frequency of g for all values in I. This approximation can be very efficiently implemented as will be detailed in Section 4. We will discuss next how the maximal frequencies of g are calculated.

3.2 Dynamic Transfer Functions

The spectral decomposition of g is obtained by applying a fast Fourier transform (FFT). We define the cut-off frequency that contains a pre-defined percentage of the signal's energy as the maximal frequency of g. If the transfer function changes arbitrarily during the rendering, the FFT has to be applied again. Although the FFT is a very efficient algorithm, this cannot be done in real-time for high-resolution transfer functions. To overcome this problem, we make use of the fact that in medical applications the transfer function is only changed in a very restricted way during user interaction.

Medical imaging workstations (e.g. Vital Vitrea, OsiriX Imaging Software) typically allow the user to change a 1D transfer function with simple mouse gestures. In order to highlight different anatomical structures (see Fig. 2) all four RGBA components of the transfer function $g(x)$ are shifted along the x-axis or multiplied by a constant factor. This behavior can also be modelled by defining the affine map

$$t \mapsto \Psi(t) = a_\Psi t + b_\Psi \tag{3}$$

and using this map to adjust the material density f before the transfer function g is applied. To determine the maximal frequency for a modified transfer function we have to analyse the function composition $g \circ \Psi \circ f$ instead of $g \circ f$. By noting that

$$(\Psi \circ f)' = (a_\Psi f + b_\Psi)' = a_\Psi f' \tag{4}$$

we can define the new estimate

$$T = \frac{1}{2v_g(I) \cdot |a_\Psi| \max |f'(x)|} \tag{5}$$

to incorporate the change of the transfer function into the approximation. This allows to approximate the maximal frequency of the modified transfer functions with minimal computational effort by using the spectral decomposition of the original transfer function. An FFT on g has to be performed only if a new transfer function is chosen, but not if the transfer function is altered in the way described above. It is important to remark that this perfectly fits the typical use case for medical visualization. A completely new transfer function is usually chosen only in the beginning of the examination after loading the dataset. At this stage the delay caused by the calculation of the FFT is negligible. The change of the transfer function during the interactive rendering can then be modelled as described above.

4 Implementation and Results

4.1 Implementation

The ray casting algorithm was implemented as a module for the visualization toolkit (VTK) using the C++ and OpenGL/GLSL programming languages. We use pre-integration and store the necessary pre-computed data in a 2D texture. In order to implement the adaptive sampling using the frequency based criterion, two additional textures are used. A 2D texture holds the maximal frequency $v_g(I)$ of g in the interval $I = [s_{\min}, s_{\max}]$. As detailed above these frequencies can be calculated in a pre-processing step by performing a spectral decomposition of g using the FFT.

The gradients $f'(x)$ are stored in an importance volume that is of the same size as the dataset. For each voxel the gradient of f is determined in each spatial direction. It is important that the adaptivity criterion estimates the maximal frequency of the whole region that is covered in the next sampling step. In order to achieve this, the gradients are filtered with a max-filter with a width of $2T_{\max}$ and the result is stored in the importance volume. Here, $T_{\max}$ denotes the maximal step size. Please note that this procedure essentially implements an isotropic sampling criterion. Further speed-up could be achieved if the direction of the ray was taken into account.

In order to estimate the step size we also need the values $s_{\min}$ and $s_{\max}$ in the region covered by the ray in the next sampling step. Although it is possible to store these values in an importance volume similar to the gradients $f'(x)$,

this approach would be very memory intensive. To circumvent this problem, we approximate f with its Taylor expansion

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) \tag{6}$$

around the sample point x_0. The upper and lower bound for $s_{\min}$ and $s_{\max}$ can then be calculated using the relations

$$s_{\min} = f(x_0) - f'(x_0)T_{\max} \tag{7}$$
$$s_{\max} = f(x_0) + f'(x_0)T_{\max} \tag{8}$$

and the gradients stored in the importance volume.

We conclude the description of the implementation by summarizing the additional computational effort that is associated with the adaptive method. The algorithm requires some additional floating point operations to calculate the interval $I = [s_{\min}, s_{\max}]$ values according to equation 8. Due to the fact that the transfer function changes during the rendering, the interval has to be shifted as described by the relation 3. Finally, the evaluation of the step size using the frequency approximation (see eqn. 5) has to be performed. Two additional texture fetches have to be carried out in order to obtain the gradient f' and the local maximal frequency of v_g. Most significantly, the adaptive method uses more texture memory to store the importance volume. Our implementation uses an importance volume that is of the same size as the original dataset. However, it has to be pointed out that it contains a lot of redundancy due to the use of the max-filter and can therefore be easily compressed.

4.2 Results

In this section the rendering performance of the algorithm for two typical medical datasets is outlined. The first one is a CT angiography of a human hand (Fig. 3) featuring a resolution of 300x300x300 voxels. The second one is a section of an abdominal CT scan with a resolution of 512x512x101 voxels (Fig. 4). The test system is equipped with a NVIDIA GeForce GTX 280 graphics card with 1GB of video memory. We use a maximal step size of $T_{\max} = 8$ and an oversampling by a factor of two as discussed in Section 2.

The regular, non-adaptive algorithm is used to render a reference solution. The transfer function is not changed during the rendering. This allows to compute the spectral decomposition of $g \circ f$ in a pre-processing step using the FFT. The solution that is obtained using the pre-computed frequencies is referred to as direct sampling. It serves as the theoretical upper bound for the average step size. As explained in Section 3, direct sampling cannot be used when the transfer function changes during the rendering. In order to evaluate the performance of the adaptive method we compare the average number of sample points per ray as well as the number of frames per second (FPS). Additionally, the number of sample points for each ray are plotted as black and white images (see Fig. 4, 3). Many sample points are used within white regions, whereas darker regions indicate that fewer samples have been taken.

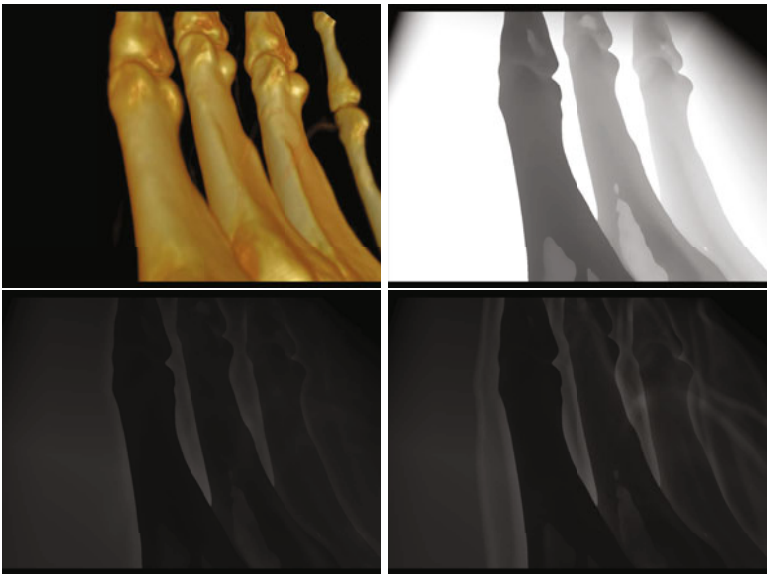


Fig. 3. Close-up view of the human hand dataset (upper left). Black and white images indicate how many samples are taken per ray if regular sampling (upper right), direct sampling (lower left) or the proposed sampling criterion (lower right) is used.

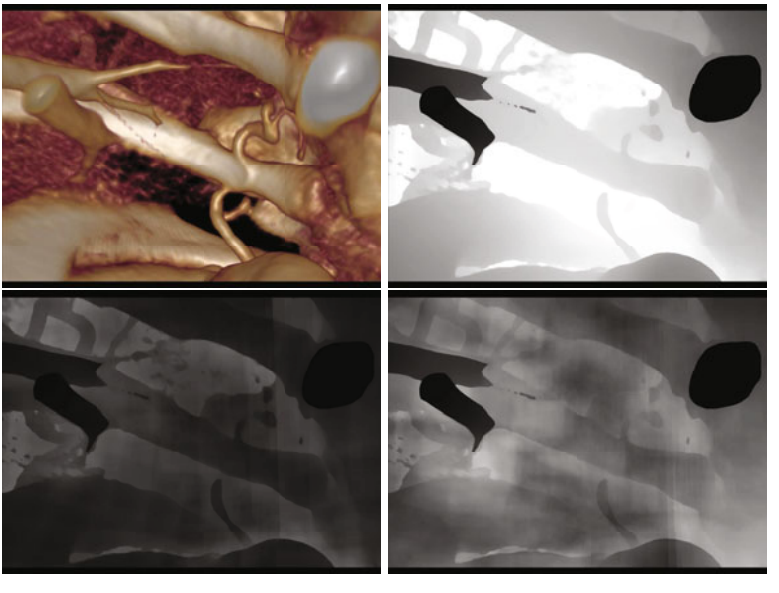


Fig. 4. Close-up view of the abdomen dataset (upper left). Black and white images indicate how many samples are taken per ray if regular sampling (upper right), direct sampling (lower left) or the proposed sampling criterion (lower right) is used.

Table 1. Performance of the proposed approximation in terms of samples per ray (less is better) and frames per second (FPS) for the hand dataset (left) and the abdomen dataset (right). Results using regular sampling and direct sampling are given as reference values.

Sampling type	regular	direct	approx.
Samples/pixel	430	75	88
FPS	1.2	6.5	5.5

Sampling type	regular	direct	approx.
Samples/pixel	752	197	404
FPS	5	8.1	6.2

No visible difference is observed between the images rendered with the different methods. The average number of sample points is reduced from 430 to 88 samples per pixel for the rendering of the hand dataset (see Table 1). This is only slightly more than the optimal solution (75 samples/pixel). A detailed analysis (see Fig. 3) reveals that most of the speed-up can be attributed to empty space skipping around the bones of the hand. In contrast to this scenario, the image generated from the abdominal CT scan does not contain much empty space. The adaptive method significantly reduces the number of sample points in all image areas which results in an average reduction of more than 50% (Fig. 4). However, the approximation still needs twice as many sample points as the optimal solution.

The adaptive method significantly increases the rendering speed. The first example shows that the acceleration factor perfectly corresponds to the reduction of the sample points. This shows that the additional computations for the adaptive method only have a very small impact. In the second example the number of samples per pixel does not scale linearly with the rendering speed which can be attributed to the parallel execution on the GPU. This suggests that an implementation wich allows a finer control over the hardware (e.g. using NVIDIA CUDA) could further accelerate the algorithm.

5 Conclusion and Outlook

We have presented a novel adaptive ray casting algorithm. In contrast to well established heuristic methods, we use the spectral decomposition of the transfer function and the dataset to derive a suitable sampling criterion. The analysis of the number of sample points per ray reveals that the new algorithm requires 50% to 80% less points compared to a non-adaptive method without any quality loss. Furthermore, the approximation of the maximal frequency of $g \circ f$ performs well compared to the optimal solution. Because of the numerical efficiency of the method, the rendering speed of the GPU is greatly increased with reference to the non-adaptive algorithm. The presented adaptivity criterion can be effectively used in a framework that uses a hierarchical data structure to perform adaptive ray casting (e.g. Gobetti et al. [4]). Such a data structure also allows to efficiently store the dataset gradients which reduces the necessary texture memory, thereby alleviating the only major disadvantage of the method. We employed the fast Fourier transform to obtain the spectral decomposition over certain regions of

g. It could be beneficial to employ a different transform that has a better spatial resolution such as a Wigner-Ville transform. Finally, the criterion could be further improved by considering anisotropic behaviour of the dataset.

Acknowledgments. We thank Frederik Giesel from the Heidelberg University Hospital for the medical datasets. This work was funded by the Klaus Tschira Foundation.

References

1. Bergner, S., Moller, T., Weiskopf, D., Muraki, D.J.: A spectral analysis of function composition and its implications for sampling in direct volume visualization. IEEE Transactions on Visualization and Computer Graphics 12(5), 1353–1360 (2006)
2. Crassin, C., Neyret, F., Lefebvre, S., Eisemann, E.: Gigavoxels: Ray-guided streaming for efficient and detailed voxel rendering. In: Proceedings of the 2009 Symposium on Interactive 3D Graphics and Games, pp. 15–22. ACM, New York (2009)
3. Engel, K., Kraus, M., Ertl, T.: High-quality pre-integrated volume rendering using hardware-accelerated pixel shading. In: Proceedings of the ACM SIGGRAPH/EUROGRAPHICS Workshop on Graphics Hardware, p. 16. ACM, New York (2001)
4. Gobbetti, E., Marton, F., Iglesias Guitián, J.A.: A single-pass GPU ray casting framework for interactive out-of-core rendering of massive volumetric datasets. The Visual Computer 24(7), 797–806 (2008)
5. Kraus, M.: Direct volume visualization of geometrically unpleasant meshes (2003)
6. Kraus, M., Strengert, M., Klein, T., Ertl, T.: Adaptive sampling in three dimensions for volume rendering on GPUs. In: 2007 6th International Asia-Pacific Symposium on Visualization, APVIS 2007, pp. 113–120 (2007)
7. Kruger, J., Westermann, R.: Acceleration techniques for GPU-based volume rendering. In: IEEE Visualization, VIS 2003, pp. 287–292 (2003)
8. Ljung, P.: Adaptive sampling in single pass, GPU-based raycasting of multiresolution volumes. In: Proceedings Eurographics/IEEE Workshop on Volume Graphics 2006, pp. 39–46 (2006)
9. Marschner, S.R., Lobb, R.J.: An evaluation of reconstruction filters for volume rendering. In: Proceedings of the Conference on Visualization 1994, pp. 100–107. IEEE Computer Society Press, Los Alamitos (1994)
10. Roettger, S., Guthe, S., Weiskopf, D., Ertl, T., Strasser, W.: Smart hardware-accelerated volume rendering. In: Proceedings of the Symposium on Data Visualisation 2003, Eurographics Association, p. 238 (2003)
11. Rubin, G.D., Beaulieu, C.F., Argiro, V., Ringl, H., Norbash, A.M., Feller, J.F., Dake, M.D., Jeffrey, R.B., Napel, S.: Perspective volume rendering of CT and MR images: applications for endoscopic imaging. Radiology 199(2), 321 (1996)
12. Schulze, J.P., Kraus, M., Lang, U., Ertl, T.: Integrating pre-integration into the shear-warp algorithm. In: Proceedings of the 2003 Eurographics/IEEE TVCG Workshop on Volume Graphics, p. 118. ACM, New York (2003)

Metrics for Uncertainty Analysis and Visualization of Diffusion Tensor Images

Fangxiang Jiao[1], Jeff M. Phillips[2], Jeroen Stinstra[3], Jens Krüger[4], Raj Varma[5], Edward Hsu[6], Julie Korenberg[7], and Chris R. Johnson[8]

[1] The Scientific Computing and Imaging Institute, University of Utah
fjiao@sci.utah.edu
[2] The School of Computing, University of Utah
jeffp@cs.utah.edu
[3] Numira Biosciences
jeroen@sci.utah.edu
[4] DFKI, MMCI, Saarbrcken
jens.krueger@dfki.de
[5] The School of Computing, University of Utah
varmanitw@gmail.com
[6] The Department of Biomedical Engineering, the University of Utah
edward.hsu@utah.edu
[7] The Brain Institute, Department of Pediatrics, University of Utah
Julie.Korenberg@hsc.utah.edu
[8] The Scientific Computing and Imaging Institute, University of Utah
crj@sci.utah.edu

Abstract. In this paper, we propose three metrics to quantify the differences between the results of diffusion tensor magnetic resonance imaging (DT-MRI) fiber tracking algorithms: the area between corresponding fibers of each bundle, the Earth Mover's Distance (EMD) between two fiber bundle volumes, and the current distance between two fiber bundle volumes. We also discuss an interactive fiber track comparison visualization toolkit we have developed based on the three proposed fiber difference metrics and have tested on six widely-used fiber tracking algorithms. To show the effectiveness and robustness of our metrics and visualization toolkit, we present results on both synthetic data and high resolution monkey brain DT-MRI data. Our toolkit can be used for testing the noise effects on fiber tracking analysis and visualization and to quantify the difference between any pair of DT-MRI techniques, compare single subjects within an image atlas.

1 Introduction

After the invention of Diffusion Tensor magnetic resonance imaging (DT-MRI) [1], a number of fiber tractography algorithms [2,3,4,5,6,7] have been proposed over the last decade. The issues of noise, motion effects or imaging artifacts create a certainty degree of uncertainty for fiber algorithms and may produce misleading tracking results. However, quantifying and effectively visualizing the accuracy and the uncertainty between results of different fiber tracking algorithms remains a significant challenge. For quantification, many fiber bundle difference metrics have been proposed [8,9], most of

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 179–190, 2010.

which use a Euclidean distance measure based upon predefined correspondences. One problem with the distance metrics is that it is easily disturbed by the predefined correspondences, with being overestimated or underestimated, as shown in Section 3. In addition, most difference metrics do not take into account the local fiber directional information and the local fiber probability information, i.e. the fraction of fibers that pass through that voxel. This will overweight the peripheral or tail voxels and ignore the directional information of the local diffusion profile. Recently, Wassermann et al. [10] put forward a Bayesian framework based on Gaussian Processes, which takes into account prior information about the fiber structure. Unfortunately, this method assumes the distribution of the fiber point position is Gaussian, which may not always to be true. In this paper we proposed three similarity metrics: the area between corresponding fiber bundles, the Earth Mover's Distance between two fiber bundle volumes, and the current distance between two fiber bundle volume that can help better quantify differences between fiber bundles and better understand uncertainty associated with fiber tracking algorithms.

Visualization of error and uncertainty is a growing area with important applications in science, engineering and medicine [11]. However, there are very few works addressing the visualization of uncertainty or the accuracy of tensor fields and specifically of fiber tracking algorithms. A recent paper by Brecheisen et al. [12], studies how to effectively visualize how the stopping criteria of FACT algorithm(Fiber Assignment by Continuous Tracking), can influence the fiber tracking results. However, this study primarily illustrates the quantification of the difference using a single algorithm and does not provide methods for inter-algorithm comparisons. Furthermore, Brecheisen et. al. use a technique in which seed points were placed manually by expert users. Such manual placement can influence the outcome of the fiber tracking algorithm and is somewhat time consuming. In this paper we describe an interactive uncertainty visualization toolkit. Users can choose different fiber tracking algorithms, change the tracking criteria, change how seed points are distributed. Furthermore, our toolkit provides the ability to track uncertainties within different anatomical regions, easily observe areas of high uncertainty and interactively explore such high uncertainty regions locally.

2 Materials and Methods

2.1 Data

Synthetic data: The synthetic data used in this paper was simulated by Numerical Fiber Generator (NFG) [13]. One B0 image ($b = 0 s.mm^2$) and twenty diffusion weighted images ($b = 3000 s.mm^2$) were obtained. The image resolution is $0.1mm \times 0.1mm \times 0.1mm$ and the image matrix size is $20 \times 20 \times 20$ voxels.

High resolution monkey brain data: The monkey brain used in this study is the right hemisphere of a whole brain. Imaging experiments were conducted on a Bruker Biospec 7-T horizontal-bore system (Bruker Inc, Billerica, MA). For data acquisition, a standard 3D diffusion-weighted spin-echo sequence was used (TR 375 ms, TE 26 ms, field of view $70 \times 51 \times 51$mm, Matrix $233 \times 170 \times 170$ which yielded an isotropic resolution of 300 microns, b-value is 2,000 s/mm^2).

Adding noise: To test the robustness of our toolkit, different levels of artificial Rician noise were added to the synthetic and the monkey brain diffusion weighted images. Six signal-to-noise (SNR) ratio levels of noise are 96,48,32,24,19 and 16, which corresponds to about 2%, 4%, 6%, 8%, 10% and 12% measured by the noise mean and divided by the signal mean. To guarantee the distribution of added noise is Rician, we proceed as follows: take the Fourier transform of the diffusion weighted image, add Gaussian noise in both the real and imaginary part of, take the magnitude of the Gaussian noise disturbed complex image, and implement the inverse Fourier transform of the magnitude image to obtain the noisy image. The same procedure was used for both synthetic data and monkey brain data. One issue that needs to be specified is that the smoothed monkey brain data was treated as the ground truth, and different levels of noise were added directly to it. This is because there is no ground truth available for real brain data and the main focus of this paper is on how to quantify and visualize the uncertainties rather than the noise issue itself.

2.2 Fiber Tracking Algorithms and Tracking Parameters

In this study, we implement six algorithms, five deterministic ones: the Streamline, Tensorline, Tensor Deflection (Tend), Guided and Fast Marching algorithm, and one probabilistic algorithm: Stochastic Tractography.

The Streamline algorithm starts from seed points and integrates along the the major eigenvector direction to form the fiber tracts. The Tensorline algorithm integrates along the following outgoing vector direction: $v_{\text{out}} = fe_1 + (1-f)((1-g)v_{\text{in}} + gD \cdot v_{\text{in}})$, which is the weighted sum of the major eigenvector direction of the current voxel e_1 and the previous voxel v_{in}, and the deflection term $D \cdot v_{\text{in}}$. Weinstein et al. [3] used a linear anisotropy measure as f, and named the technique the Tensorline algorithm. Lazar et al. [4] extended this idea to set f and g to any user defined number between 0 and 1, this is the Tend algorithm. It is worth noting that when $f = 1$, both the Tensorline algorithm and the Tensor Deflection algorithms are exactly the same as the Streamline algorithm. The Guided tracking algorithm integrates along the major eigenvector direction while being guided by *a priori* information, which can be anatomical knowledge or fiber tracking results from some other algorithms. The Fast Marching algorithm is based on a fast marching level set method where a front interface propagates in directions normal to itself with a non-negative speed function. From this speed function, three-dimensional time of arrival maps generated, which produce the connection paths among brain regions. The Stochastic fiber tracking algorithm calculates the probabilities of connections based on a Bayesian framework. To facilitate the comparisons, we use the same start and end region for all of the six algorithms. We use linear anisotropy (CL) rather than fractional anisotropy (FA) as the anisotropy value for tracking. The reason for this choice is that the tensor shape with high FA, i.e disks, do not necessarily have a clear contrast between the major and secondary eigenvalue, in which case major eigenvector direction may easily change by 90 degrees based primarily on noise effects. The step size was chosen to be 0.05 mm for the synthetic data, and 0.15 mm for the monkey brain data, while the stopping criteria was CL=0.1 for both synthetic data and monkey brain data. For all of the six algorithms, only fiber tracts starting from the seed region and ending in the end region are selected for comparison.

3 Fiber Similarity Metrics

In this section we define three distance measures between pairs of fibers A and B, as well as between fiber bundles $\mathcal{A} = \{A_1, A_2, \ldots\}$ and $\mathcal{B} = \{B_1, B_2, \ldots\}$. Each fiber is described by a sequence of points, that is fiber $A = \langle a_1, a_2, \ldots \rangle$. We can also represent a fiber A by a piecewise-linear curve defined by segments $a_i a_{i+1}$ between consecutive fiber points. More conveniently, we can just denote a set of voxels that a fiber goes through. For a fiber A, denote this set of voxels as $\bar{A} = \{\bar{a}_1, \bar{a}_2, \ldots\}$ and for a fiber bundle $\mathcal{A}$ it is denoted $\bar{\mathcal{A}} = \{\bar{a}_1, \bar{a}_2, \ldots\}$. Given a fiber bundle $\mathcal{A}$, for each voxel $\bar{a}_h$, we can then determine the fraction of fibers that pass through that voxel (the probability), denoted as $P_{\bar{a}_h}$. Additionally, we can calculate the average tangent direction of the fibers that pass through a voxel $\bar{a}_h$, denoted as $T_{\bar{a}_h}$. These quantities will be useful in the distance measures we define for comparing fibers and fiber bundles.

Before we introduce the new measures, we first comment on commonly used distance measures in the literature. Given two fibers A and B, let the *pointwise-order distance* of the common area be defined $D_{\mathsf{po}}(A,B) = \sum_{i=1} \|a_i - b_i\|$. Let B_ℓ denote the point on the piecewise-linear curve of fiber B a distance ℓ from the start by arclength, and let $\ell_A(a)$ be the distance from the start of fiber A to a point $a \in A$. Then let the *corresponding arc-length distance* be defined $D_{\mathsf{cal}}(A,B) = \sum_{i=1} \|a_i - B_{\ell_A(a_i)}\| + \sum_{j=1} \|b_j - A_{\ell_B(b_j)}\|$. Let $\phi_B(a)$ be the closest fiber point in B to point a. Then let the *corresponding closest point distance* be defined $D_{\mathsf{ccp}}(A,B) = \sum_{i=1} \|a_i - \phi_B(a_i)\| + \sum_{j=1} \|b_j - \phi_A(b_j)\|$. These measures are illustrated in Figure 1 of two fibers A and B. Although, these distances may be easy to compute, they typically take the sum or the average of distances between points, which are overestimates or underestimates of the true distances. This is due either to poor predefined correspondences, poor discretization or a complex local configuration of the fibers or fiber bundles.

For the crossing point of Fiber A and Fiber B in Figure 1, the local difference value assigned to this point for any Euclidean distance measure will be zero. Although the spatial locations of the crossing point are the same, the fiber directions at this point are different for Fiber A and Fiber B. As such, the local difference value at this point should not be zero. The area difference metric defined in Section 3.1 solves this dilemma. This local area difference metric can help to visualize the local fiber difference in a more robust way based on the spatial information. For the Earth Movers Distance and the current distance, the predefined correspondences are not needed. Therefore the problem of poor predefined correspondences, poor discretization or a complex local configuration of the fibers or fiber bundles can be successfully avoided. Furthermore, when the local fiber probability or the local fiber directional information are taken into account, this will further reduce the bias by only considering the spatial location. Thus, these two global metrics are more applicable for purpose of quantifying distances accurately.

3.1 The Area between Corresponding Fibers or Corresponding Points

We propose a distance measure $D_{\mathsf{Area}}(A,B)$ that measures the distance between two fibers A and B by the area between them. Let $\mathsf{Area}(a,b,c)$ describe the area of the triangle between points a, b, and c. Let $\psi_B(a_i)$ and $\psi_A(b_j)$ describe the mappings to points in fiber B and A, respectively, defined by the discrete Frèchet correspondence [14]; the

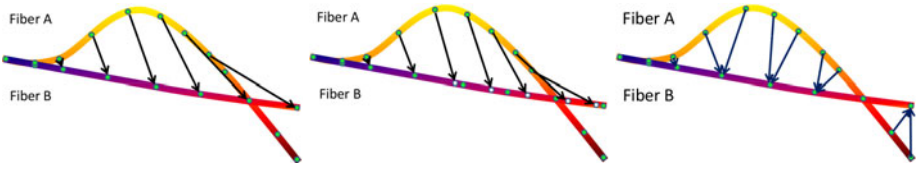


Fig. 1. Different distances: (left) $D_{\mathsf{po}}(A,B)$, (middle) $D_{\mathsf{cal}}(A,B)$, (right) $D_{\mathsf{ccp}}(A,B)$

closest distance from each point to the other fiber that also preserves the ordering along the fibers. Formally

$$D_{\mathsf{Area}}(A,B) = \sum_{i=1} \sum_{b_j, b_{j+1} \in \psi_B(a_i)} \mathsf{Area}(a_i, b_j, b_{j+1}) + \sum_{j=1} \sum_{a_i, a_{i+1} \in \psi_A(b_j)} \mathsf{Area}(b_j, a_i, a_{i+1}).$$

We can also assign a local distance measure at each point $a_i \in A$ as

$$D_{\mathsf{Area}}(a_i,B) = \frac{1}{2} \cdot [\frac{1}{2}\mathsf{Area}(a_{i-1}, a_i, \psi_B^-(a_i)) + \sum_{b_j, b_{j+1} \in \psi_B(a_i)} \mathsf{Area}(a_i, b_j, b_{j+1}) + \frac{1}{2}\mathsf{Area}(a_i, a_{i+1}, \psi_B^+(a_i))],$$

where $\psi_B^-(a_i)$ (resp. $\psi_B^+(a_i)$) is the min (resp. max) index point in $\psi_B(a_i)$. We use multiple terms for each point and divide by two so the local distance is symmetric (from A to B or B to A) and the sum or the average of local distances is the global distance.

3.2 The Earth Mover's Distance

The Earth Mover's Distance, also called Kantorovich-Wasserstein distance, can be visualized as finding the optimal way to move piles of "earth" or dirt to fill a series of holes, minimizing the total "work" or mass times distance [15]. Based on the voxelsize representation $\bar{\mathcal{A}}$ and $\bar{\mathcal{B}}$ of fiber bundles $\mathcal{A}$ and $\mathcal{B}$, the Earth Mover's Distance between two fiber bundles is defined as

$$EMD(\bar{\mathcal{A}}, \bar{\mathcal{B}}) = \frac{\sum_{i \in \bar{\mathcal{A}}} \sum_{j \in \bar{\mathcal{B}}} c_{ij} f_{ij}}{\sum_{i \in \bar{\mathcal{A}}} \sum_{j \in \bar{\mathcal{B}}} f_{ij}} = \frac{\sum_{i \in \bar{\mathcal{A}}} \sum_{j \in \bar{\mathcal{B}}} c_{ij} f_{ij}}{\sum_{j \in \bar{\mathcal{B}}} \bar{b_j}}, \tag{1}$$

where c_{ij} is the cost to move a unit of supply from $i \in \bar{\mathcal{A}}$ to $j \in \bar{\mathcal{B}}$, and f_{ij} is the flow that minimizes the overall cost

$$\sum_{i \in \bar{\mathcal{A}}} \sum_{j \in \bar{\mathcal{B}}} c_{ij} f_{ij}, \tag{2}$$

subject the following constraints:

$$f_{ij} \geq 0 \; i \in \bar{\mathcal{A}}, \; j \in \bar{\mathcal{B}}; \qquad \sum_{i \in \bar{\mathcal{A}}} f_{ij} = \bar{b_j} \; j \in \bar{\mathcal{B}}; \qquad \sum_{j \in \bar{\mathcal{B}}} f_{ij} \leq \bar{a_i} \; i \in \bar{\mathcal{A}}, \tag{3}$$

where $\bar{a}_i$ is the total supply of supplier i and $\bar{b}_j$ is the total capacity of consumer j. In this case, they both are the probability values at the ith voxel of fiber bundle $\bar{\mathcal{A}}$ and jth voxel of fiber bundle $\bar{\mathcal{B}}$. The cost function c_{ij}, which can be any predefined distance measure in any dimension, is the Euclidean distance between the fiber voxels of two fiber bundles in this paper. Therefore, the Earth Mover's Distance between two fiber bundles is the minimum effort to redistribute the probability of one fiber bundle to match the other. This measure not only takes into account the Euclidean distance but also considers the fiber probability difference as well.

3.3 The Current Distance

The current distance was proposed by Glanués and Vaillant [16] as a measure to compare a broad class of shapes (including point sets, curves, and surfaces) by how they interact with each other. Recently, Durrleman et. al. [17] investigated medical application in more depth and showed that the current distance is increasing with decreasing signal-to-noise ratio of the image. The measure can be interpreted as implicitly lifting each shape to a single point in a high (often infinite) dimensional Euclidean space, specifically, a reproducing kernel Hilbert space, where the similarity can be measured as the Euclidean distance. As such, fiber bundles can be interpreted as a set of curves, and the high dimensional vectors corresponding to each curve can be summed to create a single point representing a fiber bundle. This provides a natural distance to compare fiber bundles. Furthermore, Joshi et al. [18] showed that we can approximate the current distance between shapes arbitrarily well by a fine enough discretization. Thus, for computational reasons, we approximate each fiber A by the set of voxels $\bar{A}$ it passes through. Then the similarity between two fibers can be written as

$$\kappa(A,B) = \sum_i \sum_j K(a_i,b_j)(T_{\bar{a}_i} \cdot T_{\bar{b}_j}), \tag{4}$$

where $K(a,b)$ is a kernel function (we use the Gaussian kernel with the bandwidth h the same as the voxel size) and $(T_{\bar{a}_i} \cdot T_{\bar{b}_j})$ is the dot product between two tangent vectors. Now the current distance is defined as

$$CD(A,B) = \kappa(A,A) + \kappa(B,B) - 2\kappa(A,B). \tag{5}$$

When using a fiber bundle $\mathcal{A} = \{A_1, A_2, \ldots, A_n\}$ instead of a single fiber A_i, we can compute the similarity between two fiber bundles as

$$\kappa(\mathcal{A},\mathcal{B}) = \sum_{A_l \in \mathcal{A}} \sum_{a_i \in A_l} \sum_{B_h \in \mathcal{B}} \sum_{b_j \in B_h} K(a_i,b_j)(T_{\bar{a}_i} \cdot T_{\bar{b}_j}). \tag{6}$$

Because the similarity function κ is a summation over terms, we can accumulate the total number of fibers that pass through each voxel and take their average tangent vector in each voxel, and then we can treat each (now weighted) voxel as a single point of the fiber bundle. The self-similarity of a fiber $\kappa(A,A)$ or of a fiber bundle $\kappa(\mathcal{A},\mathcal{A})$ can be viewed as a norm of that fiber or fiber bundle, denoting how large that shape is in the high-dimensional vector space. Alternatively, the current distance between two fibers

(or fiber bundles) can be seen as the difference in how the fibers act on the underlying space, measured by how they act on each other. This action is described by its local influence in the space by the kernel function K and in the direction it flows through the tangent vector. Thus the current distance measures the difference in how two fibers (or fiber bundles) flow through a given space.

4 Results and Discussion

4.1 Fiber Track Difference Quantification

Figure 2 shows the tracking results of the Streamline, Fast Marching, Guided and the Stochastic tracking algorithm on synthetic data and on the monkey brain data. Since the Tensorline and the Tend method yield similar results to the Streamline algorithm, we only show the Streamline algorithm result. The Stochastic tracking result is embedded in each of the other three results as a semi-transparency isosurface. The colormap shows the local fractional anisotropy (FA) value. The start seed points are shown by the smaller spheres while the ending region points are shown by the larger spheres. Figure 3 shows the average closest distance (D_{ccp}) and average area between corresponding fibers of noise free volume and each level of noisy volume using four algorithm: Streamline, Tensorline, Guided and Tend algorithm, whose correspondence between fibers or points are easily defined. For the synthetic data, the tracking results from each algorithms are compared with the ground truth, and for the monkey brain data, the tracking results of each algorithms under different noise levels are compared with its own tracking result on the smoothed data without artificial Rician noise. One can see that either the average distance or the average area difference increases with the increasing noise level. The

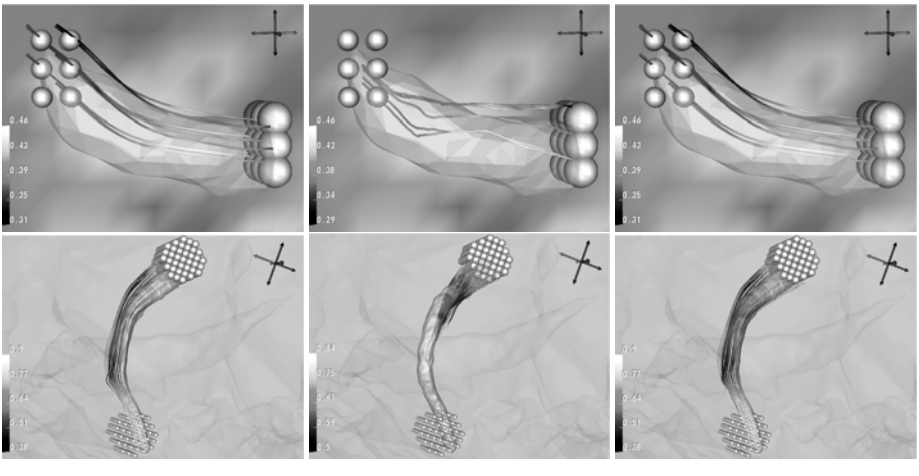


Fig. 2. The results for synthetic data (top) and monkey brain (bottom) of four tracking algorithm, Streamline (left), Fast Marching, (middle), Guided tracking (right), Stochastic tracking (embedded as isosurface), the larger sphere shows the end points, and the smaller spheres show the starting points

performance of these four algorithms are very similar, except the Guided tracking algorithm yields slightly different results from the other three methods. The fiber difference quantification using the current distance and the Earth Mover's Distance for both synthetic and monkey brain data are shown in Figure 4. The fiber tracks generated using all of the six tracking algorithms are compared with the ground truth or smoothed monkey brain data. We can see that the Stochastic tracking algorithm is very stable at different noise levels and produces the smallest difference for both measures on both data sets, while the performance of Fast Marching Method is not stable and tends to produce quite different results from the the ground truth or smoothed monkey brain data. These comparisons suggest that the Stochastic tracking algorithm is less sensitive to noise, since the noise effects are already accounted for during fiber tracking process. Furthermore, this suggests that the Stochastic fiber tracking algorithm may be good at finding the major structure of the data set, even at a very low signal to noise ratio. The Earth Mover's Distance and current distance can effectively capture the level of uncertainty for most of the algorithms, and the distances tend to increase when the noise level increase.

Although further detailed validation is required, the three metrics put forward in this study show the potential for quantifying the difference between fibers. The area difference is good at local uncertainty visualization and quantification, which we will address in the next subsection, however it needs predefined correspondence. Both the Earth Mover's Distance and the current distance are global measures, but do not need any correspondences. Therefore, the combination of these metrics can help to quantify the uncertainty or accuracy both locally and globally.

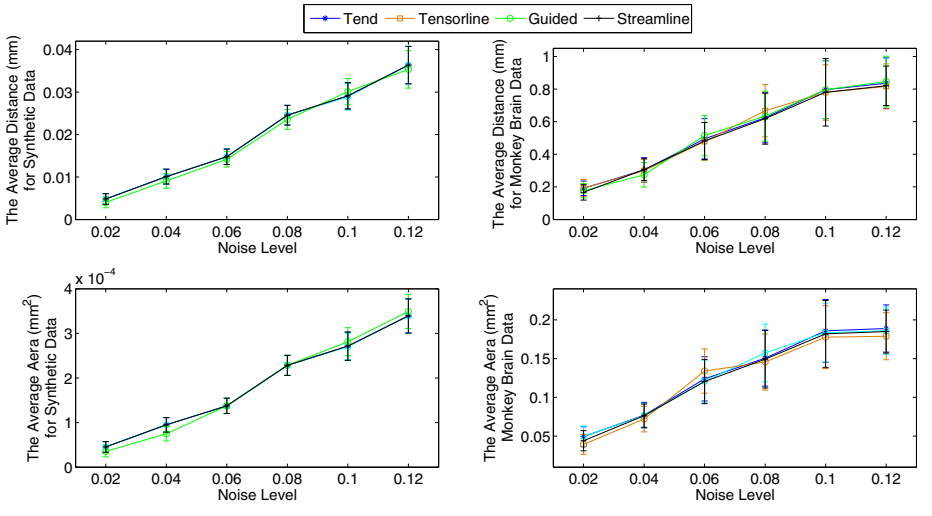


Fig. 3. The average distance (top) and average area (bottom) between fiber tracking results of the noise free volume and each level of the noisy volume for synthetic data (left) and monkey brain data (right)

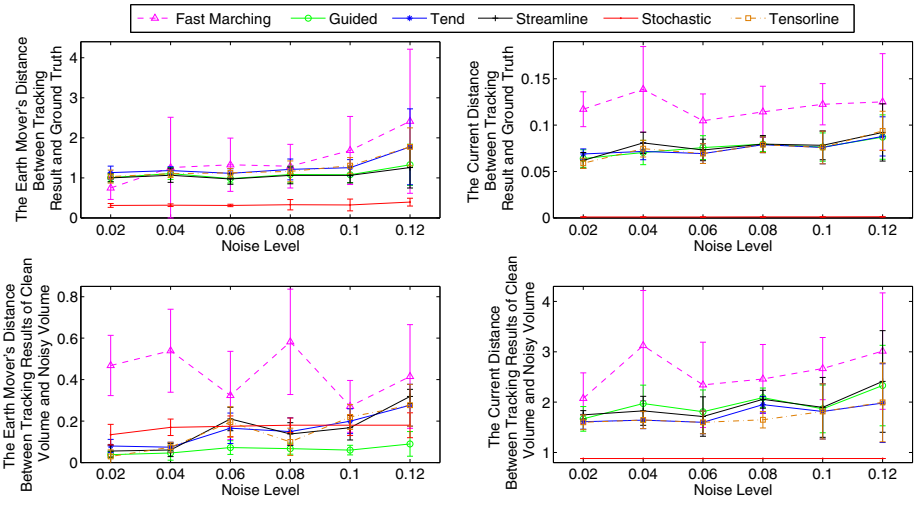


Fig. 4. The fiber difference quantification using Earth Mover's Distance (left) and current distance (right) on synthetic data(top) and monkey brain data (bottom)

4.2 DT-MRI Uncertainty Visualization Toolkit

The interactive uncertainty visualization toolkit we designed to visualize the differences between different fiber tracking algorithms, noise levels, and fiber difference metrics was created using the SCIRun problem solving environment (http://www.sci.utah.edu/software.html). After choosing two DT-MRI volumes to be compared, a user can select fiber tracking algorithms, tracking parameters such as the stopping criteria, the interpolation method and the integration method, etc. The available tracking algorithms are the six algorithms discussed previously. We note that due to computational costs, the Fast Marching and Stochastic algorithms cannot be currently used in interactive mode. The interpolation methods in the toolkit are nearest neighbor, linear, B-spline, Catmull-Rom, and Gaussian interpolation. An Euler method, as well as forth-order Runge-Kutta integration methods are used to generate the fiber tracks. The stopping criteria includes, the threshold for the length of the fiber, the local anisotropy value, the local curvature, and the number of integration steps. The user can move a widget inside the DT-MRI volume, the position of the seed points will be linearly interpolated along the widget, and the local area difference between two preselected volumes will be interactively visualized. Furthermore, the length of the widget, the shape of the widget and the seed points density can also be changed interactively. Then correspondence of fibers between any two volume is defined by whether the fibers come from the same seed points. Figure 5 illustrates the global and local visualization windows. The left hand side shows the interactive uncertainty visualization of the synthetic data, the middle column shows the interactive uncertainty visualization of the monkey brain data and the right column shows the zoom in view of the monkey brain data. The fiber tracks are generated using the Streamline algorithm. The global and local difference histograms are shown through an attached UI interface, and the local difference histogram (in red) is updated

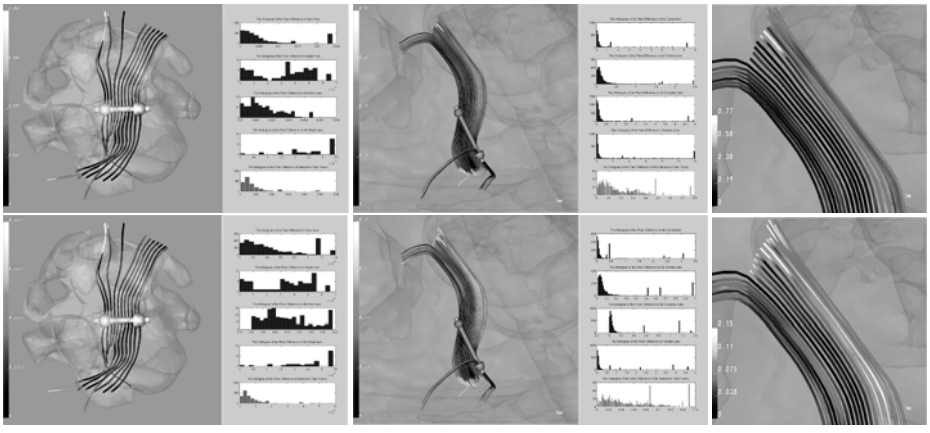


Fig. 5. The interactive visualization of local closest distance difference (top) and local area difference (bottom) of the synthetic data (left), monkey brain data (middle) and the zoom in view of the monkey brain data (right)

interactively. Through this interactive UI, the user can easily compare the uncertainty or accuracy of the current fiber track with fiber tracks from different anatomical regions, which helps quickly locate areas with high uncertainty.

In general, the end points of the fibers have a larger uncertainty due to the accumulated tracking error. As shown in Figure 5, these areas are highlighted and easily located by the average area metric rather than average closesest distance metric, especially within the monkey brain data. One can also notice that the area with high uncertainty is located to the right and towards the end of the tracking for the monkey brain. While this area is visible in the distance difference visualization, it is more clearly highlighted through the local area difference visualization upon closer inspection at the right column. Taken together, a user can interactively explore, quantify, and visualize uncertainties within DTI-MR data using the our uncertainty visualization toolkit. We note that noise is only one of many potential DTMRI uncertainty sources. Imaging artifacts, partial voluming or even different ber tracking parameters can also produce uncertainties. Although we only focus on the uncertainty associated with different levels of noise, the toolbox we developed in this study can be used as a tool to quantify and visualize any kind of uncertainty.

5 Conclusion and Future Work

In this paper, we put forward three metrics to quantify the difference between two fiber bundles. The quantification results on synthetic data and the monkey brain data show that the area between corresponding fibers can effectively capture the local or global uncertainty. The Earth Mover's Distance, which considers the local fiber probability, also shows good quantification of the fiber difference. The current distance metric, which considers the local fiber probability, the local fiber directional information illustrates

the power of quantifying the global uncertainty. Based on all of these metrics, we illustrated an interactive uncertainty visualization toolkit within the SCIRun environment that includes six fiber tracking algorithms were implemented and associated tracking parameter and noise level options. The location and the density of the seed points can be changed interactively, and most importantly, the uncertainties can be visualized interactively and quantitatively compared with the fiber tracks in different anatomical regions. Thus our toolkit facilitates DT-MRI tracking algorithm comparison, the impact of noise or other artifacts, and visual uncertainty localization.

Currently, we are working on the analysis of the fiber differences between subjects from different age groups within a human brain atlas, which will quantify the variabilities of the fiber tracking results for different age groups. In future, we will apply the metrics defined in this study to fiber clustering and segmentation, which may potentially improve fiber clustering and segmentation accuracy. Fiber bundle difference quantification can be cast as a registration problem, therefore all of the other metrics already used in image registration, such as mutual information, may be useful for fiber bundle difference quantification. Furthermore, since the metrics we presented here are easily extended, we plan to compare q-ball and other higher order fiber tracking algorithms. We are also working with a group of neurologists to obtain anatomical axon tracks within the monkey brain as to compare histological ground truth of the brain connections with the tracking results of different algorithms. Finally, our interactive quantification and visualization toolkit may potentially be used as a tool for surgical planning aiding the further improvement of validation of Diffusion Tensor imaging.

Acknowledgements

The authors want to thank Suresh Venkatasubramanian, Yarden Livnat, J. Davison de St. Germain, Osama Abdullah and Tom Close for their useful discussions.

References

1. Basser, P., Mattiello, J., Lebihan, D.: Estimation of the effective self-diffusion tensor from the nmr spin echo. Journal of Magnetic Resonance, Series B 103, 247–254 (1994)
2. Basser, P., Pajevic, S., Pierpaoli, C., Duda, J., Aldroubi, A.: In vivo fiber tractography using dt-mri data. Magnetic Resonance in Medicine 44, 625–632 (2000)
3. Weinstein, D., Kindlmann, G., Lundberg, E.: Tensorlines: Advection-diffusion based propagation through diffusion tensor fields, pp. 249–253 (1999)
4. Lazar, M., Weinstein, D., Tsuruda, J., Hasan, K., Arfanakis, K., Meyerand, M., Badie, B., Rowley, H., Haughton, V., Field, A., Alexander, A.: White matter tractography using diffusion tensor deflection. Human Brain Mapping 18, 306–321 (2003)
5. Parker, G., Wheeler-Kingshott, C., Barker, G.: Estimating distributed anatomical connectivity using fast marching methods and diffusion tensor imaging. IEEE Transactions on Medical Imaging 21, 505–512 (2002)
6. Cheng, P., Magnotta, V.A., Wu, D., Nopoulos, P., Moser, D.J., Paulsen, J., Jorge, R., Andreasen, N.C.: Evaluation of the gtract diffusion tensor tractography algorithm: A validation and reliability study. NeuroImage 31, 1075–1085 (2006)

7. Friman, O., Farnebäck, G., Westin, C.F.: A bayesian approach for stochastic white matter tractography. IEEE Transactions on Medical Imaging 25, 965–978 (2006)
8. Corouge, I., Fletcher, P.T., Joshi, S., Gouttard, S., Gerig, G.: Fiber tract-oriented statistics for quantitative diffusion tensor mri analysis. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 131–139. Springer, Heidelberg (2005)
9. Goodlett, C.: Computation of statistics for populations of diffusion tensor images. In: Computation of Statistics for Populations of Diffusion Tensor Images (2009)
10. Wassermann, D., Bloy, L., Verma, R., Deriche, R.: A gaussian process based framework for white matter fiber tracts and bundles, applications to fiber clustering. In: Proceedings Medical Image Computing and Computer Aided Intervention (MICCAI Workshop), London, pp. 200–214 (2009)
11. Johnson, C.: Top scientific visualization research problems. IEEE Transactions on Computer Graphics and Applications 24, 13–17 (2004)
12. Brecheisen, R., Vilanova, A., Platel, B., Ter Haar Romeny, B.: Parameter sensitivity visualization for dti fiber tracking. IEEE Transactions on Visualization and Computer Graphics 15, 1441–1448 (2009)
13. Close, T., Tournier, J.D., Calamante, F., Johnston, L., Mareels, I., Connelly, A.: A software tool to generate simulated white matter structures for the assessment of fibre-tracking algorithms. NeuroImage 47, 1288–1300 (2009)
14. Eiter, T., Mannila, H.: Computing discrete Frèchet distance. Technical Report CD-TR94/64, Christian Doppler Laboratory for Expert Systems, TU Vienna, Austria (1994)
15. Kantorovich, L.: On a problem of monge. Uspekhi Mat. Nauk. 3, 225–226 (1948)
16. Vaillant, M., Glaunès, J.: Surface matching via currents. In: Proc. Information Processing in Medical Imaging, vol. 19, pp. 381–392 (2005)
17. Durrleman, S.: Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution. Thèse de sciences (phd thesis), Université de Nice-Sophia Antipolis (2010)
18. Joshi, S., Kommaraju, R.V., Phillips, J.M., Venkatasubramanian, S.: Matching shapes using the current distance (Manuscript: arXiv:1001.0591)

Robust 3D Reconstruction and Mean-Shift Clustering of Motoneurons from Serial Histological Images

Nicolas Guizard[1], Pierrick Coupe[1], Nicolas Stifani[2], Stefano Stifani[2], and D. Louis Collins[1]

[1] McConnell Brain Imaging Centre, Montréal Neurological Institute, McGill University, Montréal, Québec, H3A 2B4, Canada
[2] Center for Neuronal Survival, Montréal Neurological Institute, McGill University, Montréal, Québec, H3A 2B4, Canada

Abstract. Motoneurons (MNs) are neuronal cells involved in several central nervous system (CNS) diseases. In order to develop new treatments and therapies, there is a need to understand MN organization and differentiation. Although recently developed embryo mouse models have enabled the investigation of the MN specialization process, more robust and reproducible methods are required to evaluate the topology and structure of the neuron bundles. In this article, we propose a new fully automatic approach to identify MN clusters from stained histological slices. We developed a specific workflow including inter-slice intensity normalization and slice registration for 3D volume reconstruction, which enables the segmentation, mapping and 3D visualization of MN bundles. Such tools will facilitate the understanding of MN organization, differentiation and function.

1 Introduction

Motoneurons (MNs) are neuronal cells from the central nervous system (CNS) whose axons extend outside of the CNS. Several MN diseases, such as amyotrophic lateral sclerosis (ALS), lead to MN cell death correlated with a progressive loss of muscle contractibility. MN specification relies notably on expression of a defined set of transcription factors. Each combination of transcription factors ultimately leads to the formation of a MN pool (or bundle) that shares common characteristics and projects to a common target. Typical amniote limbs are composed of more than 50 different muscles [17], however little is known about the intrinsic identity of the corresponding MN bundles. Recent studies with embryo mouse models, have characterized pool specific transcription factors such as Pea3 [8], Scip and Runx1 [7] providing basic knowledge of MN pool differentiation.

Understanding the mechanisms and the precise topography of the transcription factors underlying MN specification will provide important insights for the

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 191–199, 2010.

elaboration of regenerative therapies (*i.e.* stem cells therapy). In this field, light microscopy is commonly used to identify MN clusters on 2D embryo mouse slices, with manual identification of the clusters [4]. Usually, an expert identifies the bundle location, the number of MN cells and the cluster diameter on a slice-by-slice basis. However, this process is time consuming and lacks reproducibility when attempting to apply a robust analysis across populations in animal studies. Furthermore, the discontinuity of the 2D histological slices prevents the investigation of the 3D organization of MN clusters. Therefore, there is a need for robust automatic 3D reconstruction, segmentation and clustering.

In this article, we propose a fully automatic approach to MN cluster identification that consists of: inter-slice intensity normalization, reconstruction of corrupted slices, non-linear symmetric inter-slice registration and a 3D Mean-Shift clustering of the MNs.

2 Materials and Methods

2.1 Data Acquisition

Runx1 lacZ/+ mice were generated and genotyped as described in [11]. For embryonic staging, the day of appearance of the vaginal plug was considered as 0.5 embryological day (E0.5). All animal procedures were conducted in accordance with the guidelines of the Council for Animal Care. E13.5 mouse embryos were collected and fixed in 2% paraformaldehyde, 10 mM sodium periodate, and 70 mM l-lysine for 2 hours; transferred to 30% sucrose for 24 hours. After freezing, 14μm cryostat sections were prepared and subjected to immunohistochemistry as described in [11] in order to identify specific MNs. All the resulting slices were captured with an EXi Retiga color camera (QImaging) mounted on an Axio Imager M1 microscope (Zeiss) with a 1.68x1.68μm in-plane resolution (Fig. 1).

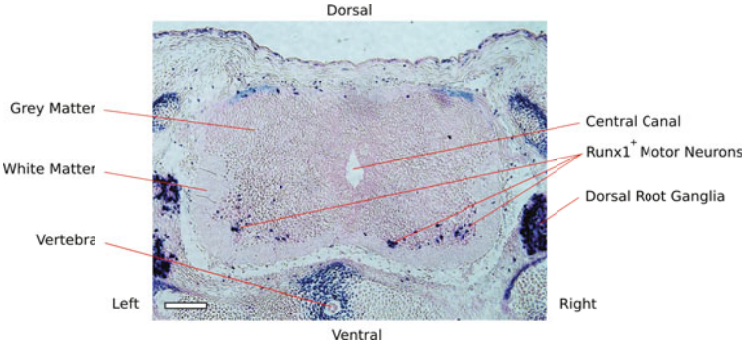


Fig. 1. Example of stained histological embryo spinal slice (scale bar = 40μm) with specific MN staining (blue).

2.2 Workflow

Identification of MN clusters requires segmentation and grouping of MN neurons on each histological slice. While MN neurons are dark and easily segmented, clustering is difficult because: i) intensity differences exist from slice to slice, ii) staining and sectioning may result in lost or corruption of slices (*e.g.* air bubbles, stretching...), and iii) the slices are not aligned within the stack. For each of these limitations we propose an adapted image processing as follows:

Inter-Slice Intensity Normalization. Staining inhomogeneity, due in part to local slice thickness variance, can produce intensity inhomogeneities between slices that adversely affect visual analysis and automated registration methods. We used the intensity normalization method proposed by [16] which consists in a linear transformation of the current slice histogram to match the histogram of the reference slice, which in this case is defined as the middle slice of the spinal cord.

Inpainting. Cryotomic image acquisition lead to discrepancies such as folding, stretching, splitting as well as air bubbles which can get stuck between the slides. In order to recover the corrupted slices identified during the image acquisition procedure, we applied a robust inpainting method [9]. Initially proposed in magnetic resonance imaging, this approach reconstructs the missing voxels by using the two most similar patches from the previous and the subsequent slices.

3D Image Reconstruction. To correct for morphological inconsistencies between the slices due to stretching, distortion, rotation and translation during the slice image acquisition, we propose an iterative 3D reconstruction method based on inter-slice registration as in [10] [5] [6]. While [10] and [5] used a Single Slice (SS) before and after to determine the deformation field, and [6] used a Multi-Slice (MS) reconstruction approach, we proposed a Multi-Slice reconstruction approach based on a symmetric registration non-linear registration and Gaussian-Distance weighted interpolation of deformation field (MSDWsym). The ANIMAL non-linear registration algorithm [2] allows for the registration parameters to be set for histological image dimensions [6]. ANIMAL uses a multi-scale vector deformation estimation with a normalized cross-correlation similarity measure. Local registration is achieved in a hierarchical manner with a Gaussian blurring of input images with kernels of varying size to recover large deformations. To enforce the inter-slice registration consistency and to reduce the effect of outlier slices with erroneous anatomy, the deformation fields T_i (Eq. 1) of the slice i is obtained with a Gaussian distance-weighted average deformation [15] to the 6 nearest slices. This proposed MSDW approach gives a stronger weight to the nearest slices such that:

$$\overline{T}(x)_i = \frac{\sum_{n=-3}^{3} \overline{T}(x)_{i \to n} \exp\left(\frac{n^2}{9}\right)}{\sum_{n=-3}^{3} \exp\left(\frac{n^2}{9}\right)} \tag{1}$$

where $\overline{T}(x)_i$ is the deformation field in x of the image i and n the distance between the reference and the target slices. This iterative process uses the previous iteration deformation field as an initial deformation for the following registration and the iteration number was set to 25. In order to produce less pair-wise registration errors, to avoid registration bias and to preserve the topology of the images [1], we force ANIMAL to be symmetric. To do so, we applied regularization constraints on the forward and the inverse deformation fields [12] for each registration of the MSDW:

$$\overline{T}_{sym}(x)_{i\to j} = \frac{\overline{T}(x)_{i\to j}.(\overline{T}(x)_{j\to i}.\overline{T}(x)_{i\to j}) + (\overline{T}(x)_{j\to i}.(\overline{T}(x)_{i\to j}.\overline{T}(x)_{j\to i}))^{-1}}{2} \tag{2}$$

where $\overline{T}(x)_{i\to j}$ is the deformation field in x of the image i to the image j.

Clustering and localization of the MNs. MNs are organized in bundles, with a tubular form along the spinal cord [14]. Due to the slice-by-slice discontinuity, manual outlining of the 2D slices cannot take in consideration the 3D structure of the MNs. A 3D reconstruction step is thus required to realign the structure within a consistent 3D volume to recover the 3D organization of the MNs. Therefore the clustering was performed on the position of the mask-outed MNs extracted from the 3D reconstructed volume using a colour filter. Because the number of MN clusters is still unknown, we applied a spatial Mean-Shift clustering algorithm [13] since it presents the advantage of being a non-a-priori technique. To achieve the Mean-Shift based clusterization, the 2D position $X(x, y)$ of the MNs was used. Moreover, to prevent the loss of continuity between the clusters due to staining and acquisition artefacts that may have omitted MNs, we propose to apply a multi-slice Mean-Shift algorithm. The MN position of the 3 slices before and after are projected into the 2D space of the current slice. The cluster centroids of the current slice are obtained by using all these projected points. This procedure is repeated for each slices of the 3D volume. Since we expect that each cluster is separated by a distance of 30-50μm [4], we set the bandwidth kernel to 40μm. Finally, during the label merging step, the cluster center found in the previous slices is used as a prior to set the label of the next one.

2.3 Experiments

Simulation: We compared 5 different 3D inter-slice reconstruction algorithms: the SS approach [5], MS approach [6] and the proposed approaches with and without the symmetrization (MSDWsym and MSDW). To validate the 3D reconstruction algorithms, we simulated on 100 identical histological slices stacked together, different random translations, rotations and shearings, which were chosen to mimic acquisition variations (Table 1). The quality of the reconstruction of each method compared was measured by estimating the MN inter-slice overlap

	Mean	Std	Min	Max
Rotation (°)	0.3	2.8	-4.7	4.9
Translation (μm)	-0.7	4.6	-8.3	8.3
Scale	1.0	0.1	0.87	1.12
Shear	0.0	1.2	-2.0	1.92

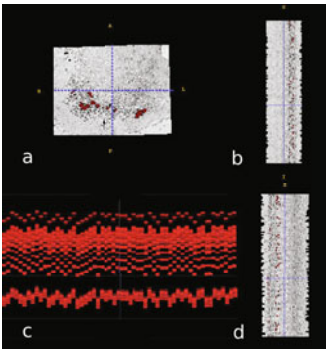


Table 1. Simulation transformation parameters and stack of the 100 identical slice after simulation: axial (a), sagittal (b), coronal (d) and 3D rendering of the MNs(c)

with the Dice's Kappa (DK) [3] agreement measure between the ground truth (Fig. 2) and each reconstruction:

$$DK = 2.V(I \cap A)/(I \cup A) \tag{3}$$

where I and A are respectively the initial aligned and automaticly aligned MNs volumes (V). DK value are comprised between 0-1 with 1 indicating perfect agreement. The topology preservation was assessed by comparing the total MN volume difference before and after reconstruction ($|V(I-M)|/V(I)$) where the value closest to 0 the value is the better the topology is preserved.

Real data: On a mouse embryo spinal cord data set of 180 slices, we first compared the histogram distance before and after the intensity normalization with the symmetric Kullback-Leibler divergence coefficient (KL). Essentially, small differences between histograms correspond to a lower divergence value. The recent 3D inter-slice reconstruction algorithm of MS [6], our proposed MSDW and MSDWsym methods were evaluated. For both methods, the final results of the clustering was assessed with manual slice-by-slice outlining performed by an expert with a DK agreement measure. The expert performed manual outlining of the clusters on 10 slices in the native space. The automatic clustering of the 10 identical slices after the 3D reconstuction was transformed back in the native space. For each of the methods, the DK agreement measure was compute between the expert and the automatic clustering into the native space.

3 Results and Discussion

3.1 Simulation

The 3D reconstructions of the simulated stack obtain by each method are presented in Fig. 2. Visual inspection of the reconstruction shows that the SS reconstruction method resulted in the worst alignment, compared to the multi-slice

Table 2. MN DK overlap measure and topology preservation with MN volume difference prior and after 3D reconstruction

Method	DK	% Volume difference
SS	0.559	1.9±2.3
MS	0.657	1.4±1.7
MSDW	0.671	1.6±2.1
MSDWsym	0.688	0.4±1.8

approaches (MS, MSDW and MSDWsym). Between the multi-slice approaches, MS and MSDW tend to a similar registration quality which seems to misalign the smaller MN clusters present on the top of the images. Because of the regularization, MSDWsym preserves the topology of small structures and provides the best overall registration. Table 2 represents the DK value of the MN mask for the 100 slices after convergence and the percentage of volume preservation. These quantitative results show that MSDWsym reaches higher DK value compared to the MS and MSDW methods. Furthermore the MN volume difference prior to and after reconstruction is best preserved using MSDWsym.

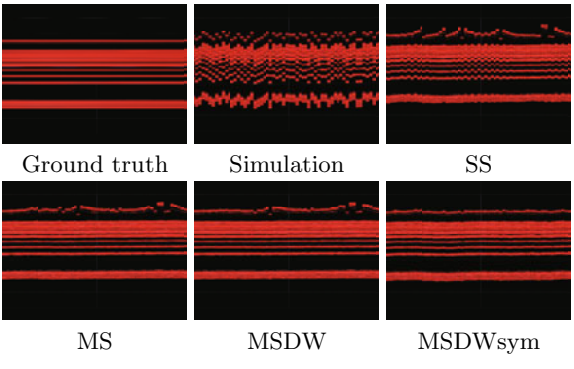


Fig. 2. Sagittal slice of the MN extracted simulated stack

3.2 Real Data

Fig. 3 shows the mouse embryo spinal cord before and after processing. Both sides of the spinal cord are processed separately and the results are visualized together. The slice-to-slice intensity inhomogeneity and mis-registration is visible in the sagittal (c, d) and the coronal (e, f) slices of Fig. 3. After processing, these artefacts are greatly reduced. The average slice-to-slice KL = 0.714 +/- 0.946 before normalization and 0.060 +/- 0.145 afterwards, indicating good agreement of the intensity histograms after normalization. The DK measure (Table 3) between the automatic and the expert segmentation for the different algorithms shows a higher agreement for the MSDWsym 3D reconstruction method. On

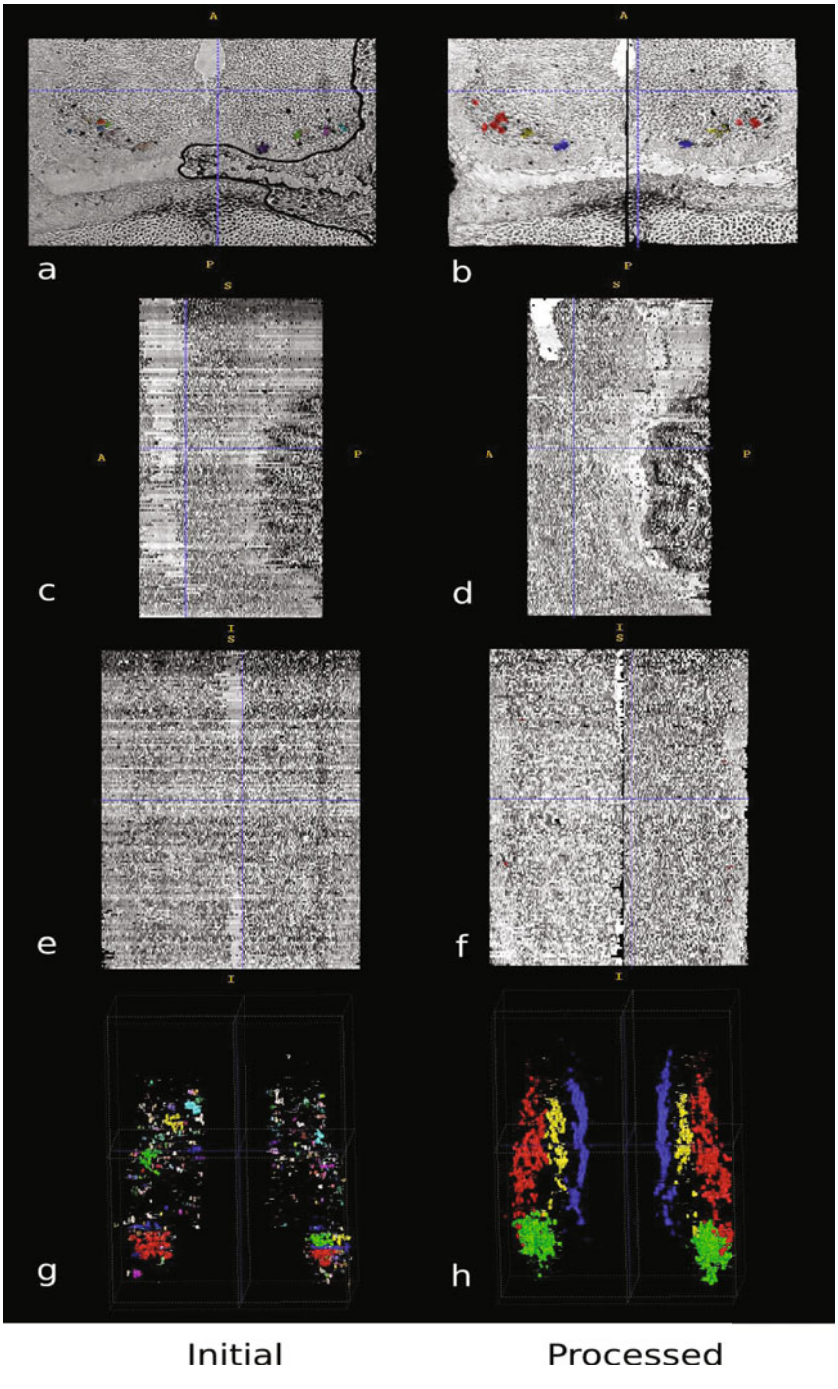


Fig. 3. MN clustering before (a,c,e,g) after (b,d,f,h) 3D reconstruction: axial (a, b), sagittal (c, d), coronal (e, f) and 3D rendering of the MN clusters (g, h)

Table 3. DK value of expert versus automatic clustering

		DK		
		MS	MSDW	MSDWsym
Cluster	#1	0.622	0.650	0.710
	#2	0.580	0.570	0.700
	#3	0.650	0.670	0.677

the axial slices (a, b) of Fig. 3, the inpainting algorithm removed the air bubble but preserved the anatomy of the slide, and most importantly preserved the MN localization and size. Compared to clustering obtained on initial stack (seen in g on Fig. 3), the MN bundles clustering obtained after reconstruction with MSDWsym algorithm (seen in h on Fig. 3) show a continuity and a spacial organization of the MNs.

4 Conclusion and Future Work

We presented a new approach to investigate MNs differentiation topology in a reproducible and automatic manner. From multiple 2D histological slices of stained mouse embryo spinal cord, we proposed a worklfow to correct for inter-slice intensity inhomogeneity, slice reconstruction, 3D volume reconstruction and MN clustering. We validated our 3D reconstruction method on synthetic data and obtained a better slice-to-slice alignment with the proposed symmetric Multi-Slice Distance-Weighted algorithm (MSDWsym). On real data, the proposed clustering method provides good agreements with the expert manual approach. These results indicate that our method can automatically identify the groups of MNs thank to a sliding Mean-Shift method and a robust 3D reconstruction. Each cluster identified on the left and right side of the spinal cord might target specific limb muscles, and we can also identify sub-clusters that might have more specific targeting. This exploratory study will enable more extensive analyses and validations in the future. In conclusion, the proposed method will facilitate the investigation of different mouse embryo populations and therefore provide important insights for future MN therapies.

References

1. Christensen, G.E., Johnson, H.J.: Consistent image registration. Medical Imaging, IEEE TMI 20(7), 568–582 (2001)
2. Collins, D.L., Evans, A.C.: Animal: Validation and applications of non-linear registration-based segmentation. IJPRAI 11, 1271–1294 (1997)
3. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology 26(3), 297–302 (1945)
4. Arber, et al.: Ets gene er81 controls the formation of functional connections between group ia sensory afferents and motor neurons. Cell 101(5), 485–498 (2000)

5. Chakravarty, et al.: The creation of a brain atlas for image guided neurosurgery using serial histological data. NeuroImage 30(2), 359–376 (2006)
6. Chakravarty, et al.: Three-dimensional reconstruction of serial histological mouse brain sections. In: ISBI, IEEE International Symposium, pp. 987–990 (2008)
7. Dasen, et al.: A hox regulatory network establishes motor neuron pool identity and target-muscle connectivity. Cell 123(3), 477–491 (2005)
8. Livet, et al.: Ets gene pea3 controls the central position and terminal arborization of specific motor neuron pools. Neuron 35(5), 877–892 (2002)
9. Manjón, et al.: Fetal mri slice reconstruction using 3d inpainting. NeuroImage 47, S72 (2009)
10. Ourselin, et al.: Reconstructing a 3d structure from serial histological sections. Image and Vision Computing 19(1-2), 25–31 (2001)
11. Stifani, et al.: Suppression of interneuron programs and maintenance of selected spinal motor neuron fates by the transcription factor aml1/runx1. International Journal of Developmental Neuroscience 8, 877 (2008)
12. Tao, et al.: Symmetric inverse consistent nonlinear registration driven by mutual information. Computer Methods and Programs in Biomedicine 95(2), 105–115 (2009)
13. Fukunaga, K., Hostetler, L.: The estimation of the gradient of a density function, with applications in pattern recognition. IEEE TMI 21(1), 32–40 (2003)
14. Jessell, T.M.: Neuronal specification in the spinal cord: inductive signals and transcriptional codes. Nature Reviews. Genetics 1(1), 20–29 (2000)
15. Lee, H., Hong, H.: Robust surface registration using a gaussian-weighted distance map in pet-ct brain images. In: Sanfeliu, A., Cortés, M.L. (eds.) CIARP 2005. LNCS, vol. 3773, pp. 794–803. Springer, Heidelberg (2005)
16. Nyúl, L.G., Udupa, J.K.: On standardizing the mr image intensity scale. Magnetic Resonance in Medicine 42(6), 1072–1081 (1999)
17. Sullivan, G.E.: Anatomy and embryology of the wing musculature of the domestic fowl (gallus). Australian Journal of Zoology 10(3), 458 (1962)

DTI Connectivity by Segmentation

Marc Niethammer[1], Alexis Boucharin[1], Christopher Zach[2], Yundi Shi[1], Eric Maltbie[1], Mar Sanchez[3], and Martin Styner[1]

[1] University of North Carolina at Chapel Hill,
[2] Swiss Federal Institute of Technology (ETH) Zurich
[3] Emory University

Abstract. This paper proposes a new method to compute connectivity information from diffusion weighted images. It is inspired by graph-based approaches to connectivity definition, but formulates the estimation problem in the continuum. In particular, it defines the connectivity through the minimum cut in tensor-weighted space. It is therefore closely related to prior work on segmentation using continuous versions of graph cuts. A numerical solution based on a staggered grid is proposed which allows for the computation of flux directly through diffusion tensors. The resulting global connectivity measure is the maximum diffusive flow supported between two regions of interest.

1 Introduction

Diffusion weighted magnetic resonance imaging (DW-MRI) can be used to measure local water diffusion in tissues in vivo. Local diffusion information is commonly used to infer connectivity information (for example in the brain) through tractography. Connectivity measures have been defined using streamline tractography, probabilistic tractography [1], optimal path approaches [2], or by geodesics in a tensor-warped space [3]. One of the major shortcomings of classical streamline tractography is its sensitivity to noise, due to the random drift occurring during the underlying integration process and due to regions with ambiguous orientation information. Probabilistic tractography alleviates the problem of orientation ambiguity by sampling large numbers of fibers, but it is still based on streamline tractography. Optimal path approaches on the other hand approach the tractography problem by finding the shortest path between two regions of interest. This is conceptually nice, because the solution space is better constrained, however, in practice these methods are prone to taking shortcuts, which requires tight control through masking to obtain anatomically meaningful connections.

Zalesky et al. [4] recently defined a connectivity measure by computing the maximum flow through a graph. They identify three desirable properties, which do not hold for standard connectivity measures: 1) independence to length of the connection, 2) proportionality to bundle cross-sectional area, and 3) invariance of the measure under exchange of source and target regions.

The approach developed in this paper is inspired by [4]. However it formulates the solution in the continuum (removing the need for graph construction)

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 200–210, 2010.

and uses the full tensor information (opposed to only the direction of the major eigenvector) for the connectivity computation. This paper is not concerned with finding the actual path, but rather the maximally supported flow between two regions of interest using a tensor-valued metric is computed. More general measures could easily be incorporated into the framework. The proposed method constitutes a continuous tensorial version of a minimum cut/maximum flow computation and is related to recent work in image segmentation [5,6].

Major advantages of the proposed method over previous work are: (i) The method allows for the easy integration of tensor information (which is much harder for a standard graph-based solution) and is extensible to more general descriptions of diffusion. (ii) The method is essentially parameter-free and directly computes a physically meaningful quantity which can be used as a surrogate measure for connectivity. (iii) The associated optimization problem can be solved reliably, because it is convex and therefore the globally optimal solution can be computed.

Sec. 2 formulates the segmentation problem. Sec. 3 discusses the numerical solution. Synthetic and real results are presented in Sec. 4. Sec. 5 concludes with an outlook on future work.

2 Segmentation/Maximum-Flow Problem

One of the simplest diffusion models is the diffusion tensor which relates local diffusion properties to measured DW signals through the Stejskal-Tanner equation

$$S_k = S_0 e^{-b g_k^T \mathcal{D} g_k},$$

where b is the b-value, g_k are the gradient directions, S_0 is the baseline image, S_k are the diffusion weighted images, and $\mathcal{D}$ denotes the diffusion tensor. $\mathcal{D}g_k$ can be interpreted as the diffusive flux according to Fick's law of diffusion. A global measure of connectivity between two regions of interest, which fulfills the properties given by Zalesky [4] can then be defined as the maximal diffusive flow between the two regions according to Fick's law.

Given a source ($\mathcal{S}$) and a target ($\mathcal{T}$) region, the maximal diffusive flow by an underlying tensor field assuming an incompressible fluid without internal sources or sinks can be obtained by solving the segmentation problem associated with minimizing the energy

$$E(u) = \int \|\mathcal{D}\nabla u\| \, dx, \;\; x \in \Omega, \;\; u \in \{0,1\}, \;\; u(x) = 1 \;\forall\; x \in \mathcal{S}; \;\; u(x) = 0 \;\forall\; x \in \mathcal{T}, \tag{1}$$

where u is an indicator function, labeling voxels between the cut and the source with 1, and 0 everywhere else, and $\Omega \subset \mathbb{R}^3$ is the domain of definition (in this setting typically given by a white matter mask), $\mathcal{S}, \mathcal{T} \subset \Omega$. The discontinuity set of the optimal solution

$$u^* = \operatorname*{argmin}_u E(u),$$

can be considered the tensor-valued, continuous equivalent of a minimum cut in a graph. The maximum diffusive flow is the flux integrated across this discontinuity set or more simply the energy value at the optimum, i.e., $E(u^*)$.

Note that an approach to tractography and connectivity information based on the diffusive flux (Fick's law) has been previously proposed by O'Donnell [3]. However, in this case the (implied) optimization problem is to minimize

$$E(u) = \frac{1}{2}\int \|\mathcal{D}^{\frac{1}{2}}\nabla u\|^2 \, dx, \;\; x \in \Omega, \;\; u \in \mathbb{R}, \;\; u(x) = 1 \;\forall\; x \in \mathcal{S}; \;\; u(x) = 0 \;\forall\; x \in \mathcal{T}.$$

This results in a smooth concentration field which can be used to define correspondences between source and target. A connectivity measures can then be defined by integrating the flux along the correspondence trajectories between source and target. In contrast, the minimization of Eq. 1 results directly in a measurement of the *maximal flow* by integrating the flux over the discontinuity set (the tensor-weighted total variation) with $u \in \{0, 1\}$. There is no need to explicitly compute the correspondences.

The energy 1 is non-convex, because the domain of u is a non-convex set. However, relaxing the condition on u to $u \in [0, 1]$ results in a convex optimization problem [6]. See also [5,7] for segmentation methods using isotropic metrics. Despite this relaxation, the solutions obtained for u are essentially binary [6], i.e., any segmentation obtained by thresholding u^* at a value $\theta \in (0, 1)$ will be globally optimal. The discontinuity set indicates the location of the minimum cut. The energy is also directly related to anisotropic total variation regularization [8].

The relaxed problem can be solved by the equivalent minimax problem

$$\{u^*, p^*\} = \arg\min_u \max_p \int p^T \mathcal{D}\nabla u \; dx, \quad \|p\| \leq 1 \quad u \in [0, 1]. \tag{2}$$

The dual energy is

$$E_{dual} = -\int_S div(\mathcal{D}^T p) \; dx + \int_{\mathring{\Omega}} \min(0, div(\mathcal{D}^T p)) \; d\mathring{\Omega},$$

where $\mathring{\Omega}$ denotes all voxels which are neither source nor target voxels. The dual energy can be used to assess the quality of a current solution iterate, since it is a tight lower bound on the energy [9].

While increased flow values may be obtained due to the inclusion of flux contributions of anatomically questionable fibers, the result of the method is (in contrast to optimal path methods) not dominated by potential shortcuts. Instead, the value of the maximum flow will be dominated by the influence of the strongest connection between source and target. Note that a major advantage of the proposed method is that it is completely parameter-free unlike standard methods based on streamline tractography, which typically have parameters controlling step-size, curvature, minimal and maximal length, integration method used, etc.

3 Numerical Considerations

The gradient descent/ascent scheme to solve 2 is

$$u_\tau = div(\mathcal{D}^T p) \quad p_\tau = \mathcal{D}\nabla u,\ \|p\| \leq 1, \quad u \in [0,1], \tag{3}$$

which is known as the Arrow-Hurwitz-Uzawa method [10]. Implementation requires the computation of the divergence of the transformed dual variables p as well as the gradient of the indicator function u. To measure flow through tensors at the location where the tensors are defined the gradient operator should be chosen such that it is evaluated colocated with p. This is achieved by a staggered grid, as depicted in Fig. 1[1]. Using trilinear interpolation (for an image volume – assuming for presentational simplicity isotropic voxels) for the indicator variables surrounding a dual variable

$$\begin{aligned} u(x+dx, y+dy, z+dz) = (1-dx)(1-dy)(1-dz)u(x,y,z) \\ + dx(1-dy)(1-dz)u(x+1,y,z) + \cdots; \quad dx,\ dy,\ dz \in [0,1], \end{aligned}$$

the gradient is simply the average of the gradients in the respective spatial directions, i.e.,

$$(u_x)_{i+\frac{1}{2},j+\frac{1}{2},k+\frac{1}{2}} = \frac{1}{4} \sum_{a,b \in \{0,1\}^2} \left(u_{i+1,j+a,k+b} - u_{i,j+a,k+b}\right),$$

and similarly for the other spatial directions. The spatial derivatives for the divergence operator need to be chosen such that it is the adjoint of the gradient operator for u, therefore

$$(p_x)_{i,j,k} = \frac{1}{4} \sum_{a,b \in \{-\frac{1}{2},\frac{1}{2}\}^2} \left(p_{i+\frac{1}{2},j+a,k+b} - p_{i-\frac{1}{2},j+a,k+b}\right),$$

and similarly for the other spatial directions, so that $div(p) = p_x + p_y + p_z$. The dual variables p are defined strictly inside the domain of the indicator variables u (as illustrated in Fig. 1). The derivatives of p at the boundary of the domain are computed such that values for p for indices outside the domain are set to zero. These are the natural boundary conditions for $-div(\cdot)$ to be adjoint to ∇ on the staggered grid.

Using a modified gradient descent for saddlepoint problems due to Popov [11] instead of the direct scheme 3 accelerates convergence. This is a fast primal-dual iteration scheme, which was used successfully by Pock et al. [12] for the minimization of the Mumford-Shah functional. The algorithm for the tensor-valued case is given in Fig. 2.

[1] When defining both the gradient and the divergence operator on the same grid, for example by forward and backward differences respectively, the solution becomes asymmetric. Further, the minimum cut will lie in-between voxels, instead of cutting through their centers. The staggered grid avoids this undesirable behavior.

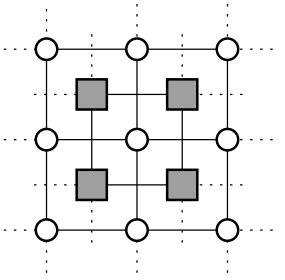


Fig. 1. Illustration of the principle of the staggered grid. Flow vectors p are defined at the location of the gray squares, indicator variables u are defined at the locations of the circles. Consequentially, gradients are defined on the squares and divergence on the circles. This allows for computation of flow through the tensors themselves.

Data: Seed points, tensor field $\mathcal{D}$
Result: Indicator function u, flow field p, maximum-flow value μ
Transfer the seedpoints from original grid to the staggered grid ;
Initialize source and target points with $u = 1$ and $u = 0$; other points to $u = \frac{1}{2}$;
Initialize $p = 0$, $u_a = u$, $u_b = u$;
repeat
  Compute gradient: $g = \nabla u_b$;
  Update flow field: $p = p + \tau \mathcal{D} g$;
  Ensure $\|p\| \leq 1$: $p = p/(\max(1, \|p\|)$;
  Compute divergence: $d = div(\mathcal{D}^T p)$;
  Update u: $u_a = u + \tau d$; $u_b = 2u_a - u$; $u = u_a$;
  Compute relative duality gap: $\Delta = \frac{\text{energy-dual energy}}{\text{energy}}$
until *convergence* $(\Delta \leq \theta)$;
μ is $\int \|\mathcal{D}\nabla u\| \, d\Omega$.

Algorithm 1: Primal-dual solution method.

Fig. 2. Algorithmic description of the primal-dual method of [12] as applied to the tensor-segmentation case

4 Results

Sec. 4.1 shows the performance of the algorithm on synthetic experiments and compares results to [4]. Sec. 4.2 applies the algorithm to a set of DT-MR images of macaques at age 2 weeks and 6 months.

4.1 Synthetic Experiments

A simple two-dimensional synthetic example for crossing fibers, illustrated in Fig. 3, shows the behavior of the algorithm under noise and directional ambiguities (due to crossings under the diffusion tensor model). The noise-free tensors were chosen as

$$T_1 = \frac{1}{1000}\begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad T_2 = \frac{1}{1000}\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$$

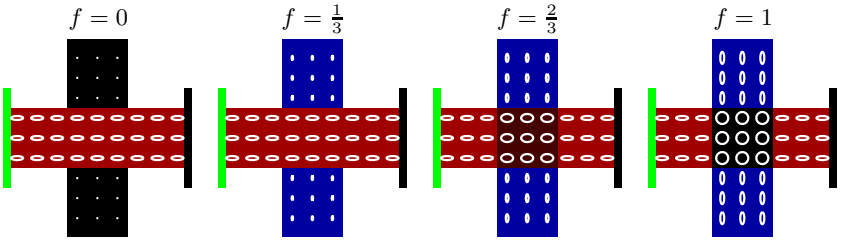


Fig. 3. Noise-free crossing case for different crossing fractions. The green and the black bars indicate the source and the target regions respectively. The red and blue colors are the planar equivalents of color-by-orientation plots, where blue denotes flow in the up/down-direction and red in the left/right direction. The intensity is modulated by fractional anisotropy of the underlying tensor.

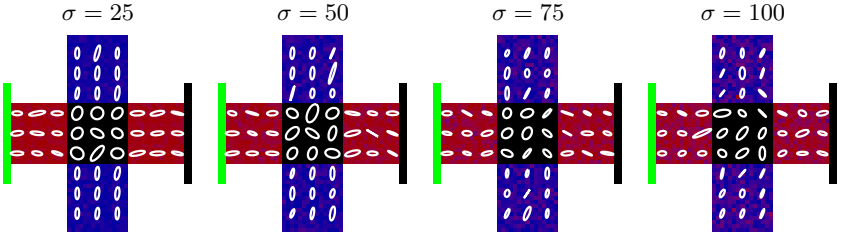


Fig. 4. Noisy crossing case, to illustrate the noise magnitudes applied

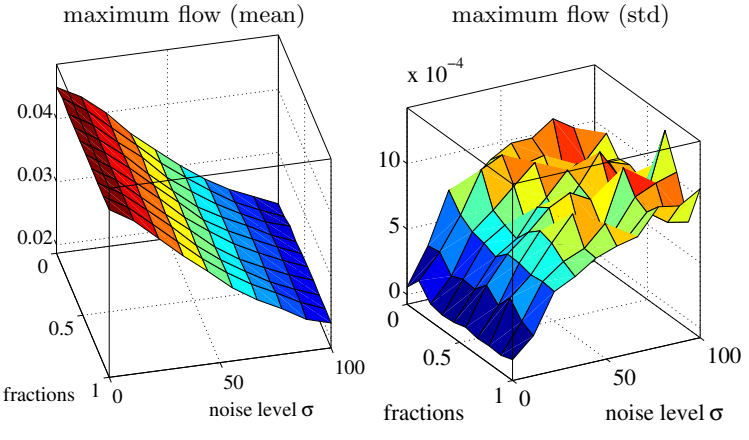


Fig. 5. Mean energy values (left) and their standard deviation (right) for 25 repetitions of 100 different σ/fraction pairings for the synthetic crossing dataset. Maximum flow decreases with increases in noise level, but is highly consistent under changes in crossing fraction.

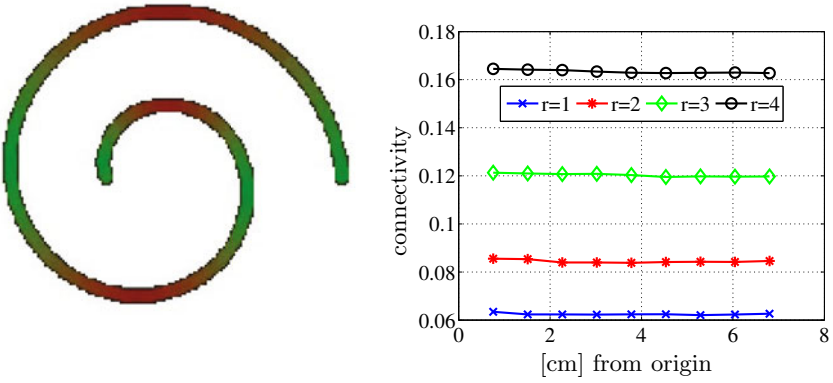


Fig. 6. Synthetic spiral as proposed in [4] (left). Estimation results for varying radii of the spiral (right; 1, 2, 3, 4 mm radius, bottom to top) for increasing distances of the target region from the source region along the spiral.

respectively. The eigenvalues of the tensors for the vertical strip were weighted by a scaling factor $f \in [0, 1]$ and tensors were combined in the overlapping region by selecting the maximum eigenvalues of the two tensors in both directions. To add noise, DWIs were computed with b-value $b = 1000$, baseline intensity $S_0 = 1000$ from the tensors at degrees $\{0,\ 22.5,\ 45,\ 67.5,\ 90\}$ (defining the two-dimensional gradient directions) in the plane. Rician noise with varying σ was subsequently added. Least-squares tensor estimation resulted in a noisy-tensor. Tensors are zero outside the crossing strips.

A set of experiments was performed combining noise levels for $\sigma \in [0, 100]$ and crossing fractions $f \in [0, 1]$. Fig. 5 shows the mean and the standard deviation for the computed maximum flow values for a set of 25 random repetitions of the experiments over the complete range for f and σ. The exact value for the noise-free experiment is $45e - 3$ (obtained by integrating the diffusive flux over a cross section), which is matched by the computed value in the noise-free case. While the computed flow value decreases with increases in the noise level (noise causes tensors to reorient, resulting in an overall decrease of diffusive flow as illustrated in Fig. 4), its value is consistent with changes in the crossing fraction. Estimation accuracy over the 25 repetitions is high, however an increase in estimation variance is observed for increases in noise level.

To compare the results to other methods in the literature, Fig. 6 shows results for the (recreated) synthetic Archimedian spiral by Zalesky and Fornito [4]. The synthetic dataset consists of noisy cigar-shaped tensors, which are aligned with the tangent of the Archimedean spiral and isotropic tensors at a given distance away from the spiral. See [4] for details on the dataset. DWIs for thirty gradient directions were generated for an image resolution of 256x256x14. No partial volume modeling (as in [4]) was used, instead the algorithm runs directly on the original (not upsampled) grid, using a mask with 4 mm radius around the center of the spiral. Fig. 6 shows maximum flow estimation results for radii of the spi-

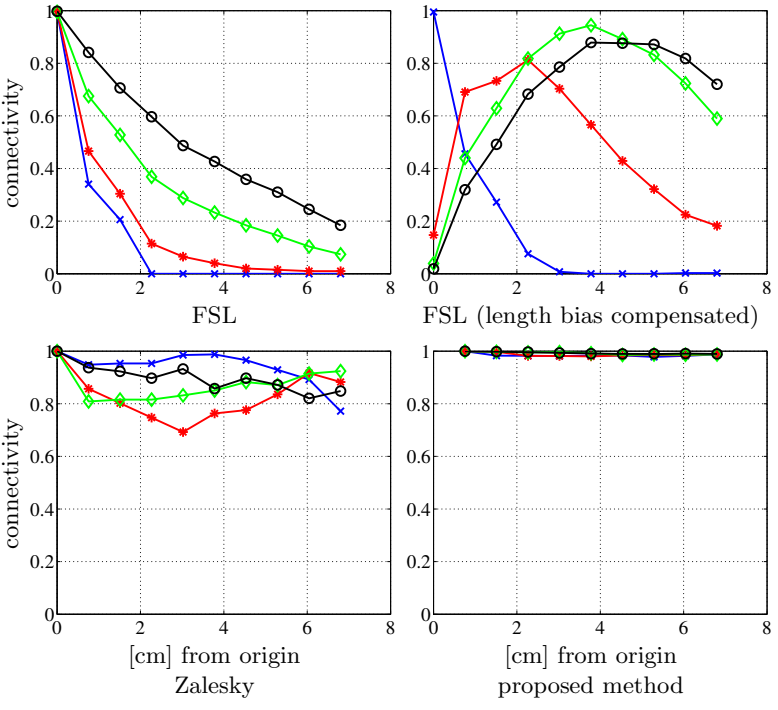


Fig. 7. Connectivity-normalized synthetic simulation results. Comparisons with FSL, FSL (length bias compensated), and Zalesky's method recreated from data of [4]. Symbols indicate spiral radius as in Fig. 6. The proposed method produces the most consistent estimation results.

ral of 1, 2, 3 and 4 mm respectively. Results represent computations between an initial source point and 9 target points distributed evenly over the spiral. The result of a good method should show increased connectivity measures for increases in radius, because this corresponds to a larger number of fibers and therefore a stronger connection. A method should also show essentially flat connectivity profiles when sampled along different points of the path, indicating invariance to the length of the path. The proposed algorithm yields highly consistent estimation results. Fig. 7 shows the normalized connectivity profiles (with a connectivity of 1 at the source location) obtained from the results in [4] for the method proposed in [4], for FSL [13], and for FSL with compensation for length bias in comparison to the maximum flow method of this paper. The maximum flow method clearly outperforms the other methods in terms of estimation consistency, obtaining a almost perfectly flat connectivity profile along the spiral.

4.2 Experiments on Real Datasets

The proposed method was applied to scans from an ongoing study of neurodevelopmental alterations caused by infant maltreatment in rhesus monkeys. Six

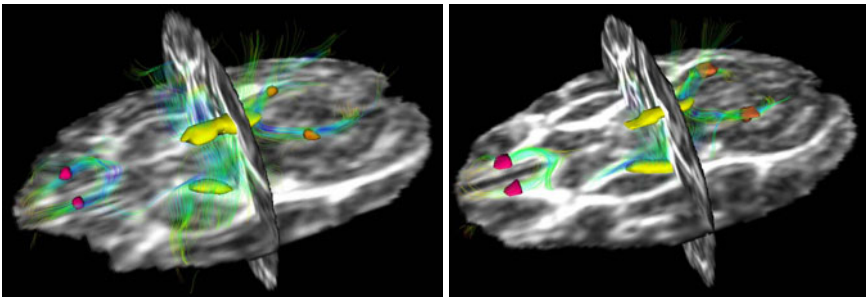


Fig. 8. ROI for the genu (pink), splenium (orange) and the internal capsule motor tracts (yellow) shown on the FA maps for the 2 weeks (left) and 6 months (right) monkey data. Fiber tracts from standard streamline method were shown to highlight the ROI, but are not used in the proposed method.

selected rhesus macaques were scanned longitudinally. Data from 2 weeks old subjects (neonates) and 6 months old subjects was analyzed. Scans were acquired on a 3T Siemens Trio scanner with 8-channel phase array trans-receiving volume coil. High-resolution T1-weighted and T2 weighted MRI scans were acquired first, followed by the DTI scans (voxel size: 1.3x1.3x1.3mm^3 with zero gap, 60 directions, TR/TE=5000/86 ms, 40 slices, FOV: 83 mm, b:0, 1000 s/mm2, 12 averages). These datasets were selected as they have SNR values at low (at 2 weeks) and intermediate (6 months) level.

An expert rater, trained in streamline fiber tractography, determined source and target regions that produced valid DTI fiber tracts for the corpus callosum genus, the splenium tracts, and for the internal capsule motor tract (see Fig. 8 for an illustration). Corresponding source and target regions were used to compute the proposed connectivity values (see Fig. 9). Connectivity computations were restricted to an expert-generated white matter mask, obtained by FA thresholding. From Fig. 9 several conclusions can be drawn: a) as expected the internal capsule has higher connectivity values than the splenium and the genu, b) splenium and genu have more similar values, but the genu is slightly higher (again expected), c) with the exception of the splenium a connectivity increase from 2 weeks to 6 months is visible, d) surprisingly the inter-subject variability is higher than the longitudinal changes.

Tract	2 weeks		6 months	
Name	Mean	Stdev	Mean	Stdev
Genu	0.051	0.008	0.057	0.003
Splenium	0.040	0.015	0.040	0.020
Internal Capsule	0.125	0.026	0.136	0.027

Fig. 9. Global connectivity values (mean and standard deviation) for 3 selected fiber tract regions in rhesus macaques at age 2 weeks and 6 months

In summary, while the tests indicated limitations of the method mainly with respect to inter-subject variability (for the dataset used), the results also point to the potential use of the maximal flow connectivity measure as an alternative to existing, tractography based connectivity measures.

5 Conclusion and Future Work

This paper described a global method to compute connectivity measures from DTI. It can be viewed as a continuous version of the approach by Zalesky [4], but additionally makes use of a tensor's diffusion strength. While the method is described in terms of the tensor model, it can be generalized to more general descriptions of diffusion. Unlike many other methods, the computed results are not based on the explicit computation of streamlines. However, a flow-field is constructed internally (given by the dual variable). Since this flow field is not unique it is not a replacement for streamline tractography at this stage. However, the computed maximum flow value is globally optimal (identical for all optimal flow fields), can be reliably obtained due to the underlying convex optimization problem of the method, and can be used as a surrogate measure for connectivity. An interesting future research direction is to construct admissible flow fields which adhere as much as possible to the underlying tensor-field, while preserving the flow magnitude and direction at the location of the minimum cut. This would allow for the reconstruction of flow fields (and fibers) through ambiguous regions.

Acknowledgements. This material is based upon work supported by the National Science Foundation (NSF) under Grant No. (EECS-0925875), by the National Institutes of Health (NIH) under Grant Nos. (R21 HD55255-2, U54 EB005149-01, RC1AA019211-01, P50 MH078105, P50 MH078105-01A2S1, R42 NS059095-03) and by UNC IDDRC HD 03110. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of NSF or NIH.

References

1. Behrens, T.E.J., Woolrich, M.W., Jenkinson, M., Johansen-Berg, H., Nunes, R.G., Clare, S., Matthews, P.M., Brady, J.M., Smith, S.M.: Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magnetic Resonance in Medicine 50, 1077–1088 (2003)
2. Lenglet, C., Prados, E., Pons, J., Deriche, R., Faugeras, O.: Brain connectivity mapping using Riemannian geometry, control theory and PDEs. SIAM Journal on Imaging Sciences 2(2), 285–322 (2009)
3. O'Donnell, L., Haker, S., Westin, C.F.: New approaches to estimation of white matter connectivity in diffusion tensor MRI: Elliptic PDEs and geodesics in tensor-warped space. In: Dohi, T., Kikinis, R. (eds.) MICCAI 2002. LNCS, vol. 2488, pp. 459–466. Springer, Heidelberg (2002)
4. Zalesky, A., Fornito, A.: A DTI-derived measure of cortico-cortical connectivity. IEEE Transactions on Medical Imaging 28(7), 1023–1036 (2009)

5. Appleton, B., Talbot, H.: Globally minimal surfaces by continuous maximal flows. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(1), 106–118 (2006)
6. Zach, C., Niethammer, M., Frahm, J.M.: Continuous maximal flows and Wulff shapes: Application to MRFs. In: CVPR, pp. 1911–1918 (2009)
7. Bresson, X., Esedoglu, S., Vandergheynst, P., Thiran, J.P., Osher, S.: Fast global minimization of the active contour/snake model. Journal of Mathematical Imaging and Vision 28, 151–167 (2007)
8. Werlberger, M., Trobin, W., Pock, T., Wedel, A., Cremers, D., Bischof, H.: Anisotropic Huber-L1 optical flow. In: BMVC (2009)
9. Boyd, S., Vandenberghe, L.: Convex Optimization, Cambridge (2004)
10. Arrow, K.J., Hurwicz, L., Uzawa, H.: Studies in linear and non-linear programming. Stanford University Press, Stanford (1958)
11. Popov, L.: A modification of the Arrow-Hurwicz method for search of saddle points. Mathematical Notes 28(5), 845–848 (1980)
12. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: An algorithm for minimizing the mumford-shah functional. In: ICCV (2009)
13. Behrens, T.E.J., Johansen-Berg, H., Jbabdi, S., Rushworth, M.F.S., Woolrich, M.W.: Probabilistic diffusion tractography with multiple fiber orientations. what can we gain? NeuroImage 34(1), 144–155 (2007)

Locally Weighted Regression for Estimating and Smoothing ODF Field Simultaneously

Xiaozheng Liu[1,2], Guang Yang[1], Bradley S. Peterson[2], and Dongrong Xu[1,2,*]

[1] Key Laboratory of Brain Functional Genomics, Ministry of Education, China; & Shanghai Key Laboratory of Brain Functional Genomics, East China Normal University, Shanghai Key Laboratory of Magnetic Resonance, East China Normal University, Shanghai 20062, China
[2] MRI Unit, Columbia University Dept of Psychiatry, & New York State Psychiatric Institute, NYSPI Unit 74, 1051 Riverside Drive, New York, NY 10032, USA
dx2103@columbia.edu

Abstract. High angular resolution diffusion imaging (HARDI) has become an important tool for resolving neural architecture in regions with complex patterns of fiber crossing. A popular method for estimating the diffusion orientation distribution function (ODF) employs a least square (LS) approach by modeling the raw HARDI data on a spherical harmonic basis. We propose herein a novel approach for reconstruction of ODF fields from raw HARDI data that combines into one step the smoothing of raw HARDI data and the estimation of ODF field using correlated information in a local neighborhood. Based on the most popular method of least square for estimating ODF, we incorporated into it local weights that are determined by a special weighting function, making it a locally weighted linear least square method (LWLLS). The method thus can efficiently perform the smoothing of HARDI data and estimating the ODF field simultaneously. We evaluated the effectiveness of this method using both simulated and real-world HARDI data.

1 Introduction

Diffusion Magnetic Resonance (MR) Imaging is a recent MR modality that has made possible to characterize in vivo the white matter architecture in the brain [1]. Diffusion Tensor (DT) is the most prevailing model in use for characterizing the displacement probability of water molecules. This model bases on the covariance matrix of a hypothesis of water diffusion of a zero-mean Gaussian distribution function, which is a second-order positive symmetric tensor, and supposes that the water molecules diffuse preferentially along the principal direction of the tensor. However, Gaussian probability distribution function is a unidirectional model, and thus is unable to characterize diffusion along more than one direction (e.g., in regions of fiber crossing), which is a common case in brains.

Recent advances have attempted to generalize the DT model [1] to a higher resolution acquisition technique, i.e., high angular resolution diffusion imaging (HARDI) [2], to address this issue. By sampling the q-space on shells with fixed radii

* Corresponding author.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 211–218, 2010.

and recovering the angular structure of diffusion in lieu of the 3D spin displacement probability function, HARDI attempts to characterize the highly complex fiber structure of fiber crossing in the brain.

Several approaches for reconstructing fiber orientations from HARDI data have been proposed, such as the q-ball imaging (QBI) technique [3], the Persistent Angular Structure (PAS) technique [4], the fiber orientation distribution (FOD) [5,6], diffusion orientation transform (DOT) [7] and Orientation Probability Density Function (OPDF) [8], for improved estimation of the true probability distribution of spatial displacement of water molecules, and subsequently a better inference of underlying fiber orientations. Among these techniques, QBI uses the Funk-Radon transform to estimate ODFs [4][8-13]. ODFs have also been approximated with different basis functions such as spherical harmonics [8-12], the Poussin kernel [13] and Spherical Ridgelets [14]. In general, methods based on spherical harmonic expansion [8-12] are preferred for its mathematically simplicity. Spherical Fourier analysis [9,11] is a recently developed technique of spherical harmony family for further improving the visibility of ODF maxima when the QBI model is used. In these works, spherical coefficients approximate HARDI data in the least-square sense by truncated series of spherical harmonics (SH), and the corresponding ODFs can then be conveniently recovered using the Funk-Hecke formula [9,11]. However, heavy noise in HARDI may sometimes cause negative values in the estimated ODF, a situation that is physically nonsense. A more recent method [6] was therefore developed to eliminate such negative values by minimizing a nonnegative least-squares cost function.

All these approaches work directly on HARDI data with an implicitly hypothesis that the HARDI data are good enough for ODF estimation. This unfortunately is not exactly correct. Therefore, a treatment for non-negativity or smoothing is desired for the HARDI data before ODF estimation [15,16]. There have been separated works on this issue [21,22], making the ODF estimation a two-step work: data regularization and then ODF estimation. By incorporating a regularization term and nonnegativity constraints into a cost function, some work has tried to unify the two steps into one [16].

We present a novel one-step approach to ODF estimation, local weighted linear least square (LWLLS) [17,18] based on least square (LS) methods [9,11]. This ODF estimation unifies into it a systematical and unanimous smoothing mechanism that employs the correlated information of the HARDI data in its neighborhood. This smoothing term takes into account the HARDI data along all the gradient directions together, it thus can smooth the data and estimate the ODF simultaneously in one unified step. This method is not only more efficient but also more accurate in theory because it treats the HARDI data systematically and unanimously. Compared with the ones in literature [15,16], our method does not require initial values that are usually tricky to set and ours is computationally efficient without using repetitions for iterative optimization. We evaluated the effectiveness of this proposed method using both simulated and real-world HARDI data.

2 Methods

2.1 Least Square Method with Spherical Harmonics

Because the HARDI data are real and symmetric in 3D space, a series of modified SH basis is defined as follows for expansion of the HARDI data [9].

$Y_j = \sqrt{2}\,\mathrm{Re}(Y_k^{|m|})$ *if* $-k \le m < 0$; Y_k^0 *if* $m = 0$; $\sqrt{2}\,\mathrm{Im}(Y_k^{|m|})$ *if* $0 < m \le k$

Where $Y_l^m(\theta,\phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^m(\cos\theta)e^{im\phi}, \theta \in [0,\pi], \phi \in [0,2\pi]$, P_l^m is a Legendre polynomial, and Re(·) and Im(·) are the real and imaginary parts, respectively. The (modified) SH basis [9] of order l contains $R = \frac{(l+1)(l+2)}{2}$ terms defined for $j(k,m) = \frac{k^2+k+2}{2} + m, k = 0,2,4,\ldots,l$ and $m = -k,\ldots,0,\ldots,k$. Let (θ_n,ϕ_n) be the *n*-th gradient direction of the N gradient directions and $s(\theta_n,\phi_n)$ be the HARDI signal acquired at each of the *N* gradient directions, $(\theta_n,\phi_n), n = 1,\ldots,N$. The HARDI signal vector *S* can be approximated by minimizing the Euclidean error norm:

$$\min_c \left\| S - BC \right\|^2 \tag{1}$$

where *S* is a $N\times 1$ vector, $S = [s(\theta_1,\phi_1)\ldots s(\theta_N,\phi_N)]^T$. C is the unknown $R \times 1$ vector of SH coefficients that parameterize the signal *S*. **B** is the $N\times R$ SH basis matrix whose *n*-th row is given as $\mathbf{B}_n = [Y_1(\theta_n, \phi_n)\ldots Y_R(\theta_n, \phi_n)]$. By Laplace–Beltrami matrix operator, a regularization term is added [9]:

$$\min_c \left\| S - BC \right\|^2 + \lambda C^{\mathsf{T}} LC \tag{2}$$

Where **L** is the $R \times R$ diagonal Laplace-Beltrami eigenvalues matrix with $\mathbf{L}_{jj} = -l_j(l_j+1)$, and l_j is the order of the *j*-th term. After getting the SH coefficients C, the ODF $\psi(\theta, \phi)$ can be reconstructed by, $\psi(\theta, \phi) = \mathbf{PCB}$, in which **P** is the $R\times R$ diagonal Funk-Radon transform matrix, and $\mathbf{P}_{jj} = 2\pi P l_j(0)$ and $P_{lj}(0)$ are the Legendre polynomial of degree l_j at 0.

2.2 Locally Weighted Least Square Method with Spherical Harmonics

Taking into account the local dependency of HARDI data in least squares method, LWLLS has a good smoothing effect in regression process [17,18]. We extend Eq. (1) to a locally-weighted least square method (to simplify the expression, we take 2D case with a 3x3 neighborhood as an example in the following text):

$$\min_C \sum_{k=1}^{N} w_k^{i-h,j-h} \left\| S_k^{i-h,j-h} - B_k^{i-h,j-h} C \right\|_2^2 \qquad h\text{=-1,0,1} \tag{3}$$

Where *i* and *j* are the shoulder marks denoting the current voxel under processing, and *h* for navigating all the voxels in the neighborhood. Our method will conduct regression in the neighborhood based on individually different weights for different voxels while weights for the same voxel along all diffusion directions always remain the same $w_1^{i-h,j-h} = \cdots = w_n^{i-h,j-h} = \cdots = w_N^{i-h,j-h}$.

By seeking the minima of Eq. (3) on C, we calculate C as follows

$$\mathrm{C} = (B^{\mathrm{T}} \sum^{-1} B)^{-1} (B^{\mathrm{T}} \sum^{-1}) S \tag{4}$$

$$\sum^{-1} = \begin{bmatrix} w^{i-1,j-1} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & w^{i-1,j-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & w^{i+1,j+1} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & w^{i+1,j+1} \end{bmatrix} B = \begin{bmatrix} (Y_1(\theta_1,\phi_1) \cdots \cdots Y_R(\theta_1,\phi_1))_1^{(i-1,j-1)} \\ \vdots \vdots \vdots \quad \vdots \quad \vdots \quad \vdots \vdots \\ (Y_1(\theta_N,\phi_N) \cdots \cdots Y_R(\theta_N,\phi_N))_N^{(i-1,j-1)} \\ \vdots \vdots \vdots \quad \vdots \quad \vdots \quad \vdots \vdots \\ (Y_1(\theta_1,\phi_1) \cdots \cdots Y_R(\theta_1,\phi_1))_1^{(i+1,j+1)} \\ \vdots \vdots \vdots \quad \vdots \quad \vdots \quad \vdots \vdots \\ (Y_1(\theta_N,\phi_N) \cdots \cdots Y_R(\theta_N,\phi_N))_N^{(i+1,j+1)} \end{bmatrix} S = \begin{vmatrix} s_1^{(i-1,j-1)} \\ \vdots \\ s_N^{(i-1,j-1)} \\ \vdots \\ s_1^{(i+1,j+1)} \\ \vdots \\ s_N^{(i+1,j+1)} \end{vmatrix} \tag{5}$$

Actually, Eq. (4) is a weighted linear least square defined within a given neighborhood. Suppose the neighborhood window contains M voxels, S now becomes a vector of $MN \times 1$ whose elements are $S_i^{i-h,j-h}$, and $\mathbf{B}$ becomes a $MN \times R$ matrix whose rows are $\mathbf{B}_n^{i\text{-}h,j\text{-}h} = \mathbf{B}_{N+n}^{i\text{-}h,j\text{-}h} = \ldots = \mathbf{B}_{MN+n}^{i\text{-}h,j\text{-}h} = [Y_1(\theta_n, \phi_n) \ldots Y_R(\theta_n, \phi_n)]$, C is still a $R \times 1$ vector, and $\sum^{-1}$ is a diagonal $MN \times MN$ matrix whose elements are the weighing factors $w^{i-h,j-h}$.

By differentiating Eq. (3) on SH coefficients C, we may find that Eq. (3) is basically a self-adaptive, local linear filter, known as the Nadaraya–Watson estimator (NWE) [17,18]. In particular, this estimator has the following form:

$$BC = \sum_{h=-1}^{1} w^{i-h,j-h} S^{i-h,j-h} / \sum_{h=-1}^{1} w^{i-h,j-h} \tag{6}$$

Both Eqs. (4) and (6) are obtained by differentiating Eq. (3) on SH coefficients C, but are expressed in different forms. Comparing with Eq. (6), we can see that using Eq. (4) for estimating the ODF basically is equal to combining into one the two tasks of smoothing the HARDI data $S^{i,j}$ (central voxel in the neighborhood) and then estimating the ODF using LS method. The novelty of LWLLS lies in that we use one simple linear regression procedure to perform smoothing the HARDI data and estimating ODF simultaneously.

2.3 The Weighting Function

The most commonly adapted weighting function that reflects the correlation in a neighborhood in image processing is Gaussian Kernel, which weighs based on normal distribution. The weights are usually calculated using one single measure or value, for example, distance. For better representing correlated information in neighborhood, we consider using a nonlinear bilateral filter, originally developed by Tomasi and Manduchi [19]. Such a filter is capable of simultaneously considering the Euclidean distance of both imaging intensity (HARDI data) and the spatial relationship of two voxels. The bilateral filer that we employ as a weighting function in Eqs. (3)-(6) is defined as:

$$w^{i,j} = \exp(-\frac{d^2(V_i,V_j)}{\sigma_d^2})\exp(-\frac{d^2(X_i,X_j)}{\sigma_r^2}) \quad (7)$$

Where $d^2(V_i,V_j)$ is the Euclidean distance between the physical locations of voxel V_i and voxel V_j. $d^2(X_i, X_j)$ is the Euclidean distance between $\log(X_i)$ and $\log(X_j)$; and X_i and X_j are $N \times 1$ vectors of HARDI data, where N is number of the gradient directions. σ_d is the geometric spread in the domain and σ_r is the photometric spread in the image range.

3 Experimental Results and Validation

We performed two sets of experiments to examine the effectiveness of our proposed model for optimized tensor estimation. We compared the performance of our method LWLLS of Eq. (4) with the LS method of Eq. (2) . In our experiments, 1) we simply set σ_d equal to noise derivation σ_n and σ_r equal to $10*\sigma_n$; 2) we set regularization weight λ=0.006 in Eq. (2).

In the first experiment, we used a synthesized ODF field [9,16] as shown in Fig. 1(a). The voxels in the 1*st* and 3*rd* quadrants contained 1 fiber, the 2*nd* quadrant 2 crossing fibers and the 4*th* quadrant 3 crossing fibers (with 1 fiber pointing out of the plane). Each ODF was generated using the multi-tensor method [9] with 81 gradient directions on the hemisphere (the second-order tessellation of the icosahedron). Noisy HARDI signal was generated by adding complex Gaussian noise with zero mean and standard deviation $\sigma = S0/\text{SNR}$ [9], where S0 is the baseline signal (S0=100) and SNR is the signal-to-noise ratio of S0. We computed the mean square error (MSE) $dist(\psi_t,\psi_e) = \sum_{i=1}^{i=Z}(\psi_t - \psi_e)^2 / Z$ between the estimated ODF field and true ODF field, with Z being the total number of ODFs in the field, ψ_t the true ODF field, and ψ_e the estimated ODF field. We first examined whether our proposed method offered any advantage over the conventional LS method at various levels of noise on a 6-th order SH basis. We then compared the LS and LWLLS methods using different SH orders with at a representative noise level of SNR=10 again by calculating the MSE. These experiments each based on 100 trials showed that the LWLLS method not only combined successfully the two tasks into one unified step, but also efficiently outperforms the conventional LS method consistently at various noise levels with an improvement of about 3% (Table 1), and at various SH orders with improvements from about 3% to 9% (Table 2). Moreover, the ODF of our LWLLS method is superior to that of the LS method in that our ODF characterizes the spatial orientation much more distinctively and explicitly than does the ODF of LS method (Fig. 1(e)&(f)).

In the second experiment, we applied our proposed method to a HARDI dataset acquired from a pig brain [20] (Fig. 2). HARDI data were obtained using the following imaging parameters: the diffusion-weighting gradient strength G = 61 mT/ m; pulse onset separation DELTA = 30ms; pulse width delta = 23ms; echo time (TE) = 60ms; b-value = 3146 s mm^{-2}; number of excitation (NEX)=2. The protocol also included three acquisitions with no diffusion weighting and 61 volumes with evenly distributed gradient directions. The image volume contained 10 slices with an in-plane resolution of 128x128, and a voxel size of 0.5x0.5x0.5 mm^3.

Table 1. MSE of varying noise levels

SNR	LS	LWLLS	Imprmnt
5	10.8922	10.5679	2.98%
15	10.9487	10.6414	2.81%
25	10.8258	10.5282	2.75%
35	10.9328	10.6185	2.87%

Table 2. MSE of varying SH orders

Order	LS	LWLLS	Improvement
4	15.7044	14.3010	8.94%
6	15.9202	14.8092	6.98%
8	15.9501	15.2011	4.70%
10	15.9580	15.5979	2.26%

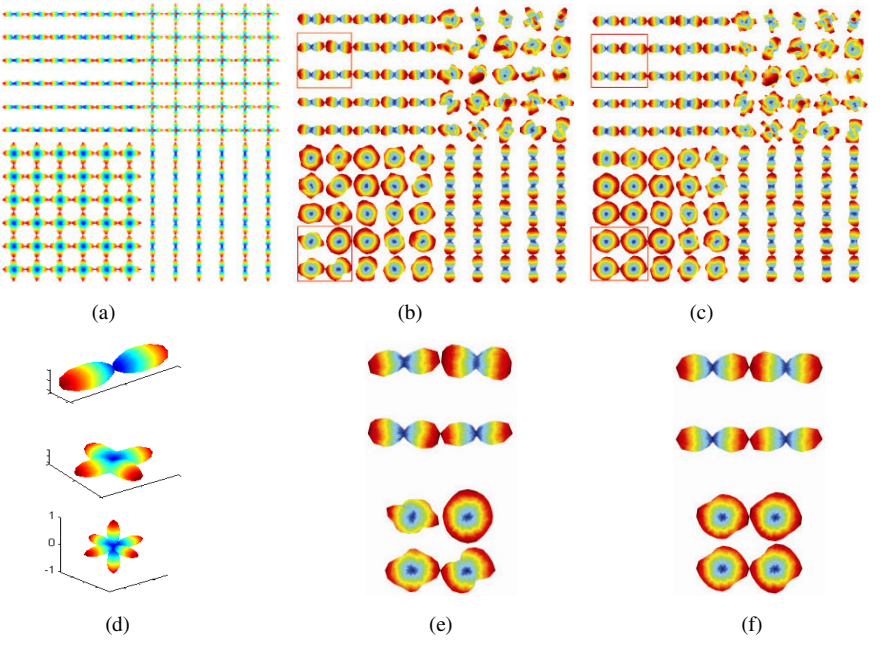


Fig. 1. The synthetic data for our first experiment. (a) The true ODF field; (b)The ODF field estimated using LS; (c) The ODF field estimated using LWLLS; (d) The profile of one single ODF corresponding to (a); (e) An amplified view of the content in the red box in (b); (f) An amplified view corresponding to the red box in (c).

Obviously, our one step framework of LWLLS works very well in terms of both smoothing and optimization of the ODF (Fig. 2). Comparable to the work in [9] and [11], the ODF profile estimated by the LWLLS is notably much sharper in denoting the underlying fiber orientations than is the ODF profile estimated by the LS, which is a desired feature of ODF for facilitating accurate fiber tracking in areas of fiber crossing.

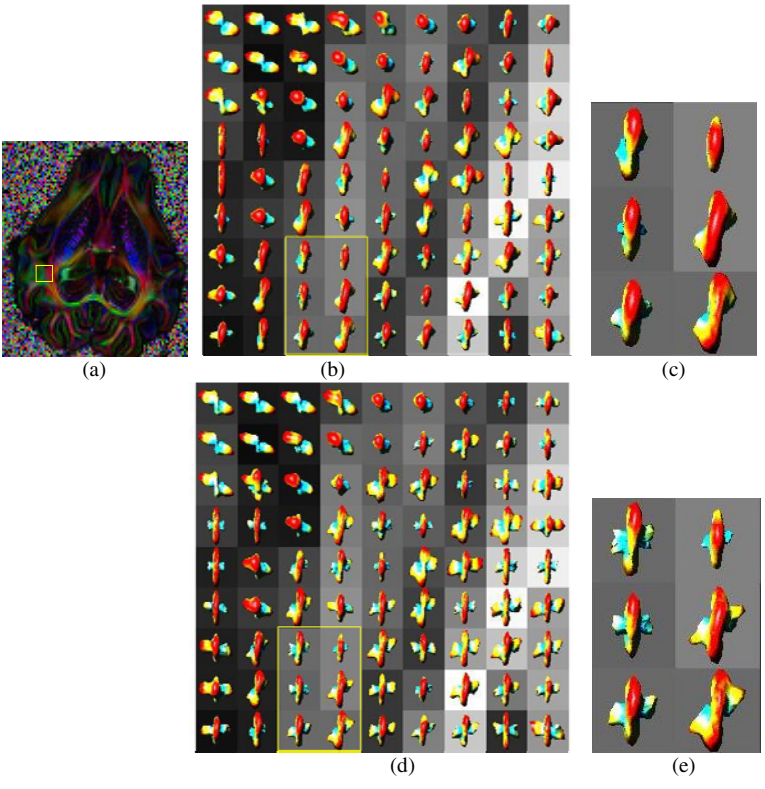


Fig. 2. Experiment results of using imaging data from a real pig brain. (a) The color fractional anisotropy (CFA) image of an axial view of the brain (Red for vertical, green for horizontal and blue for perpendicular to the view); (b) The ODF field estimated using LS corresponding to the selected area (a); (c) Amplified view of the yellow box in (b); (d) ODF field estimated using LWLLS corresponding to the selected area in (a); (e) Amplified view of the yellow box in (d). All the ODFs in (b)-(e) are superimposed on their corresponding FA image.

4 Conclusion

We have presented a novel ODF estimation method that can estimate ODF field and smooth ODF field in one step through incorporating spatial information into least square method. Experiment results using both synthetic and real imaging data demonstrated the success and effectiveness of our proposed framework.

Acknowledgement. NASARD Grant CU52051501, NIH Grant RO1MH09170, A Grant from SIMONS Foundation NIDA Grant DA017820, NIBIB Grant 1R03EB008235-01A1, NIMH Grant 5R01MH082255-03, NIEHS Grant 1 R01 ES01557901 A2/ A2S109. Cina Natinal 985 Program, Shanghai Commission of Science and Technology Grant # 10440710200, Shanghai Key Laboratory of MR, Open Project Grant Support , Two Grants from East China Normal University School of Psychology & Cognitive Neuroscience.

References

1. Basser, P., Mattiello, J., Le Bihan, D.: Estimation of the effective self-diffusion tensor from the NMR spin echo. Journal of Magnetic Resonance B 103, 247–254 (1994)
2. Tuch, D.S.: High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. Magnetic Resonance in Medicine 48, 577–582 (2002)
3. Tuch, D.S.: Q-ball imaging. Magnetic Resonance in Medicine 52(6), 1358–1372 (2004)
4. Jansons, K.M., Alexander, D.C.: Persistent angular structure: new insights from diffusion magnetic resonance imaging data. Inverse Probl. 19, 1031–1046 (2003)
5. Tournier, J.D., Calamante, F., Connelly, A.: Robust determination of the fibre orientation distribution in diffusion MRI. NeuroImage 35(4), 1459–1472 (2007)
6. Jian, B., Vemuri, B.: A unified computational framework for deconvolution to reconstruct multiple fibers from diffusion weighted MRI. IEEE TMI 26(11), 1464–1471 (2007)
7. Ozarslan, E., Shepherd, T., Vemuri, B.C., Blackband, S., Mareci, T.H.: Resolution of complex tissue micro architecture using the diffusion orientation transform (DOT). NeuroImage 31, 1086–1103 (2006)
8. Tristán-Vega, A., Westin, C.-F., Aja-Fernández, S.: Estimation of fiber orientation probability density functions in high angular resolution diffusion imaging. NeuroImage 47(2), 638–650 (2009)
9. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Regularized, fast and robust analytical Q-ball imaging. Magnetic Resonance in Medicine 58, 497–510 (2007)
10. Frank, L.R.: Characterization of anisotropy in high angular resolution diffusion-weighted MRI. Magnetic Resonance in Medicine 47(6), 1083–1099 (2002)
11. Hess, C.P., Mukherjee, P., Han, E.T., Xu, D., Vigneron, D.B.: Q-ball reconstruction of multimodal fiber orientations using the spherical harmonic basis. MRM 56(1) (2006)
12. Aganj, I., Lenglet, C., Sapiro, G.: ODF reconstruction in Q-ball imaging with solid angle consideration. In: IEEE Int. Symposium on Biomedical Imaging (2009)
13. Rathi, Y., Michailovich, O., Bouix, S., Shenton, M.: Orientation distribution estimation for Q-ball imaging. In: MMBIA (2008)
14. Michailovich, O., Rathi, Y.: On approximation of orientation distributions by means of spherical ridgelets. IEEE TIP 19(2), 461–477 (2010)
15. Assemlal, H.E., Tschumperl'e, D., Brun, L.: Robust variational estimation of PDF functions from Diffusion MR signal. In: CDMRI (2008)
16. Goh, A., Lenglet, C., Thompson, P., Vidal, R.: Estimating Orientation Distribution Functions with Probability Density Constraints and Spatial Regularity. In: Yang, G.-Z., et al. (eds.) MICCAI 2009, Part I. LNCS, vol. 5762, pp. 877–885. Springer, Heidelberg (2009)
17. Nadaraya, E.A.: On estimating regression Theory. Probab. Appl. 9(1), 141–142 (1964)
18. Ruppert, D., Wand, M.P.: Multivariate locally weighted least squares regression. Ann. Statist. 22(3), 1346–1370 (1994)
19. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: IEEE International Conference on Computer Vision (ICCV), pp. 839–846 (1998)
20. Dyrby, T., Baaré, W., Alexander, D.C., Jelsing, J., Garde, E., Søgaard, L.V.: An ex vivo imaging pipeline for producing high-quality and high-resolution diffusion-weighted imaging datasets. Human Brain Mapping (2010), doi:10.1002/hbm.21043
21. Tristán-Vega, A., Aja-Fernández, S.: DWI filtering using joint information for DTI and HARDI. Med. Image. Anal. 14(2), 205–218 (2010)
22. Descoteaux, M., Wiest-Daesslé, N., Prima, S., Barillot, C., Deriche, R.: Impact of Rician Adapted Non-Local Means Filtering on HARDI. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 122–130. Springer, Heidelberg (2008)

Distinguishing Left or Right Temporal Lobe Epilepsy from Controls Using Fractional Anisotropy Asymmetry Analysis

Hai Li, Zhong Xue*, Mario F. Dulay, Amit Verma, Solomon Wong, Christof Karmonik, Robert Grossman, and Stephen T. Wong

The Center for Bioengineering and Informatics and Methodist Neurological Institute, The Methodist Hospital, The Methodist Hospital Research Institute, Weill Cornell Medical College, Houston, TX
zxue@tmhs.org

Abstract. This paper presents an automatic fractional anisotropy (FA) asymmetry analysis and applies it to determine the FA asymmetry (FAA) changes associated with the sides of seizure origin of patients with temporal lobe epilepsy (TLE) using diffusion tensor imaging (DTI). All the control and patient images are first normalized onto the JHU-DTI-MNI atlas using a simultaneous deformable DTI registration algorithm, and the FA images are warped accordingly. Then, the tract-based spatial statistics (TBSS) algorithm is employed to quantify the FA on white matter (WM) skeletons, which are divided into different sections by overlapping with the 102 WM regions defined by the atlas. The FAA values, i.e., the relative differences of each WM skeleton section pair, are calculated. Statistical analysis is then performed to identify the regions that significantly contributed to the group differences between control and left/right TLE, as well as between left and right TLE. The results indicate that FAA correlates with the side of seizure origin, and those of certain regions are significantly different between normal controls and left or right TLE. The quantitative results can be useful for pre-surgical evaluation of TLE patients and for better understanding of the relationship between fiber tracts with the site of origin of TLE, EEG tests, the syndromes and neural psychological responses.

Keywords: Diffusion tensor imaging, temporal lobe epilepsy, tract-based spatial statistics, fractional anisotropy asymmetry.

1 Introduction

Diffusion tensor imaging (DTI) has been widely used in clinics and research for assessment of white matter in the development, pathology and degeneration of neurological diseases, surgical planning of neural surgeries, and connectivity of neural networks. Tractography provides visualization of white matter (WM) fibers in vivo and is an intuitive and meaningful approach to study the connectivity of white matter fibers. By extracting the WM bundles, we can perform quantitative analysis by

* Corresponding author.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 219–227, 2010.

associating them with measures of the fiber bundles, such as fractional anisotropy (FA) and apparent diffusion coefficient (ADC) along each bundle. All these results illustrated that the automatic and robust calculation of these fiber bundles plays an important role in order to perform accurate quantitative analysis.

Fibers can be automatically classified into different bundles in quantitative tract-based analysis. In [1], the protocols to manually reconstruct eleven major WM tracts based on region of interest (ROI) are described; however due to intra- and inter-rater variability, partial volume effects, fiber crossing and fiber branching, such ROI-based method is not robust enough to ensure all the fibers of interest are extracted. Much work has been done for an automatic neural bundle classification, such as spectral clustering [2], normalized cut [3], expectation-maximization (EM) [4], and K-means [5]. These methods focused on the automatic grouping of a large number of fiber tracts into different classes, and these methods have employed different fiber similarity measures and different classification strategies. For the individual subject, they provide nice color-coded visualization, but for fiber bundles across subjects, image registration is a necessary step to automatically label each bundle according to the anatomical structures of an atlas [6]. Because of anatomical variability among individuals and registration errors, the automatic bundle labeling tools cannot guarantee accurate extraction of all the fibers belonging to the same bundle. Therefore, fiber classification errors, registration errors and inter-subject variability are important challenges for quantitative analysis.

The tract-based spatial statistics (TBSS) is proposed in [7], which uses nonlinear image transformation and combines the strength of both voxel-wise and tractography-based analysis. In this algorithm, first, a non-linear deformable registration is used to align the subject images with a template image as accurate as possible. Then an alignment-robust tract representation called the FA skeleton is extracted, and the FA values of each subject are projected onto the skeleton. The advantage of TBSS is that compared to the registration of the entire brain image, the correspondence of the FA skeleton across different subject is much better, hence the FA analysis based on the skeleton improves the sensitivity, objectivity and interpretability of multi-subject diffusion imaging studies. In [8], TBSS is applied to investigate diffusion measures between alcohol-dependent individuals (ALC) and non- or light-drinking controls (LD), and the results showed more robust statistical results.

This paper presents an automatic FA asymmetry (FAA) algorithm. First, we apply the simultaneous diffusion image registration algorithm to align all the subject images with the template image. Experimental results showed that this algorithm generated relatively accurate results as compared with the free-form deformable registration [9]. Then, TBSS is employed to quantify the FA on the WM skeletons. Because the template has been manually segmented into 102 WM regions, we divided the FA skeleton into different sections according to the atlas. Thus, the FAA on each WM skeleton section can be calculated between the left-right skeleton section pairs. Statistical tests can then be performed for quantitative analysis.

In experiments, we applied the FAA algorithm to identify the regions contributing significant differences between control and left Temporal Lobe Epilepsy (TLE), control and right TLE, as well as left and right TLE. Statistical results of 14 normal controls, 9 left TLE and 9 right TLE patients demonstrated that FA asymmetry correlates with the side of seizure origin, and the FA asymmetry of certain regions is found to be significantly different between normal controls and left or right TLE.

2 Method

2.1 Overview

The framework of the FAA method is illustrated in Fig. 1. First, all of the input DTI data from TLE patients and normal controls are registered with the template image using the local Fast Marching (FM)-based simultaneous deformable registration [9]. 102 WM regions have been previously defined on the template image (the JHU-DTI-MNI atlas). Then, after warping all the FA maps into the template space, TBSS is employed to quantify the FA values on the WM skeletons overlaid on the 102 WM regions. The FA asymmetry on the WM skeleton section of each WM label between the corresponding left-right WM region pair is calculated. Finally, statistical analysis is performed to identify the regions contributing significant differences between control and left TLE or right TLE, respectively. In the following two sections, we briefly introduce the simultaneous diffusion image registration algorithm and FAA analysis.

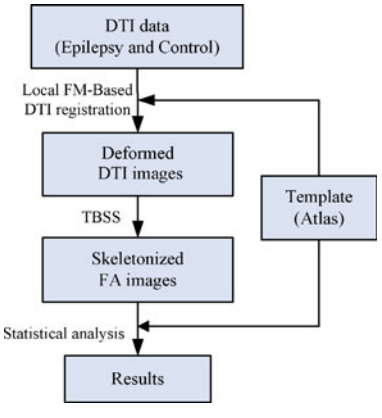


Fig. 1. The framework of the FAA method

2.2 Local FM-Based DTI Registration

Spatially aligned DTI is a key step to quantitatively compare images obtained from different subjects or the same subject at different timepoints. (An important feature for DTI registration is that tensor orientation should be considered?). Existing methods, such as the piece-wise affine transformation [10] and the diffeomorphic non-linear registration algorithms [11], use analytical gradients of the registration objective functions by simultaneously considering the reorientation and deformation of tensors during the registration. However, similar to image-intensity based registration, only voxel-wise tensor-similarity is utilized, and such a feature is relatively limited to local image information. In order to register DTI more accurately, this paper utilizes the local FM-based DTI registration algorithm proposed in [9]. That algorithm not only considers the orientation of tensors during registration but also

appropriates the neighborhood tensor information of each voxel to drive the deformation, and such neighborhood tensor information is extracted from a local fast marching algorithm around the voxels of interest [12]. The formulation of the DTI registration algorithm is similar to that of the common image registration framework. The objective function is defined as,

$$E(\mathbf{f}) = E_{sim}(D_T, D_S, \mathbf{f}) + \lambda E_{con}(\mathbf{f}), \tag{1}$$

where D_T is the template image, and D_S is the subject image. $\mathbf{f}$ is the deformation field to align the subject image. The first term is the similarity measure between the two images based on the current deformation, $\mathbf{f}$, and the second term represents the smoothness constraint of the deformation field. λ is the weight of the constraint energy function.

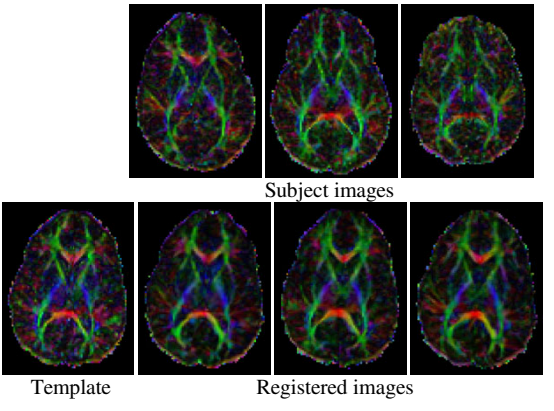


Fig. 2. Examples of the registration results of real DTI images

The cubic B-Spline registration model is used for the deformation field, and hence the smoothness of the deformation is guaranteed, and the second term can be omitted [13]. To ensure that the topology of the deformation fields is correct, the topology regularization algorithm is used so that the voxel-wise Jacobian determinants of the deformation field are positive. The detailed objective function can be re-written as,

$$E = \sum_{\mathbf{x}\in\Omega} \| A[D_S(\mathbf{f}(\mathbf{x}) + \mathbf{x})] - A[Q(\mathbf{f}(\mathbf{x}))D_T(\mathbf{x})Q^T(\mathbf{f}(\mathbf{x}))] \|^2, \tag{2}$$

where $\mathbf{x}$ is a voxel in the template image domain Ω. $D_T(\mathbf{x})$ denotes the tensor at voxel $\mathbf{x}$ in the template image, and $D_S(\mathbf{f}(\mathbf{x}) + \mathbf{x})$ represents the tensor at the corresponding location in the subject image. $Q(\mathbf{f}(\mathbf{x}))$ is the tensor rotation matrix calculated from the deformation field at voxel $\mathbf{x}$. $A[]$ represents the feature extraction operator of the diffusion tensors centered on a voxel; $\| A[D_1] - A[D_2] \|^2$ calculates the distance between two sets of such feature vectors. $A[]$ is defined by the arrival time maps of the evolving fronts through performing a local tensor-based fast

marching [9]. Although the tensors within a neighborhood of a voxel are high dimensional systems, a simple fast marching starting from that voxel can extract the tensor patterns around it. Thus, the main idea is to not only use the neighborhood tensor information but also use local fast marching time maps as tensor features for simultaneous DTI registration.

Experiments on both simulated and real DTI images are performed to validate the accuracy of the registration algorithm, respectively. The first row of Fig. 2 shows some subject images, and the second row gives the template and the registered images. Quantitative measures of the registration errors show that for simulated images, the average difference between the registration results and the ground truth is about 3.3*mm*, for images with voxelsize 2*mm*×2*mm*×2.7*mm*.

2.3 The Fractional Anisotropy Asymmetry Method

Because of the alignment inaccuracy and anatomical variability among subjects, quantitative comparison of FA values within each fiber bundle or ROI can result in inaccurate results. The TBSS method attempts to bring together the strengths of voxel-based and bundle-based methods and yields relatively stable quantitative measures. The key idea of TBSS is to extract the FA skeleton after image registration, so that the FA values determined along the FA skeletons of different subjects are more stable [7]. This is because the FA skeleton extracts lines in the centers of fiber bundles and they are generally consistent across subject. Therefore, the FA values found along the perpendicular directions of the skeleton could be stable values to capture the fiber information. In this paper, after performing image registration as described in Section 2.2, the modified TBSS method is applied to extract the FA values of each subject as shown in Fig. 3:

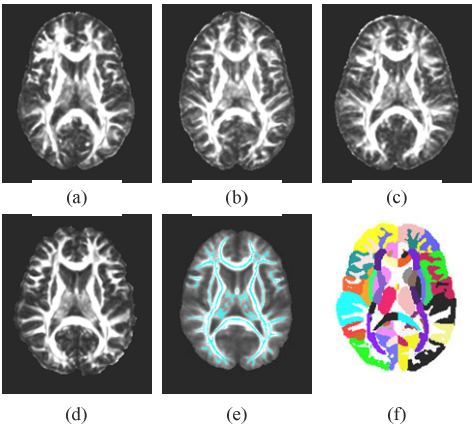


Fig. 3. Examples of FAA results. (a)-(c) Three samples of the FA images; (d) the average FA image; (e) the WM skeleton extracted from (d); (f) the tissue labels on the template image space. The skeleton is then divided into different sections according to the atlas.

1) Calculate the mean FA map from all of the warped FA images, and apply a thinning procedure to produce a skeletonized mean FA image. The skeleton is also thresholded so that it does not cover the regions with low mean FA or high inter-subject FA variability. This part is the same as the TBSS algorithm [7].
2) The skeleton is overlaid with the atlas and is divided into different sections according to the anatomical regions. By searching perpendicularly within a range of each skeleton point, the maximal FA values are projected onto the skeleton so each skeleton voxel is now assigned an FA value.
3) Finally, the FA feature of each region is defined as the average of the FA values of the corresponding skeleton section.

It can be seen that the anatomical structures are combined with TBSS for detailed quantitative analysis. Then we perform statistical analysis on the FA values along the WM skeleton of each region of interest. For each subject, the average FA values along the WM skeleton of 102 regions are calculated as:

$$\mathrm{fa}_{l,s} = \frac{1}{N(l,s)} \sum_{r \in l} \mathrm{fa}_{r,s}, \tag{3}$$

where $fa_{r,s}$ is the FA value of region (label) r of subject s belong to the same group, and $N(l,s)$ is the number of voxels for the skeleton of region l. Fig. 4 shows some examples of the mean FA skeletons on the subject image space. It can be seen that compared to the entire image, the skeletons overlap with each subject better.

Because the labels of the brain are symmetric, suppose region l_{left} and region l_{right} are the corresponding region pairs, FAA value can be calculated as:

$$\boldsymbol{a}_{l_{\mathrm{left}},l_{\mathrm{right}},s} = \frac{\mathrm{fa}_{l_{\mathrm{left}},s} - \mathrm{fa}_{l_{\mathrm{right}},s}}{\mathrm{fa}_{l_{\mathrm{right}},s}}. \tag{4}$$

The FAA values reflect the asymmetry of each anatomical region. In this work, we performed statistical tests on the FAA values among normal controls, left TLE, and right TLE. The goal is to quantitatively analyze FA changes associated with the side of seizure origin of patients with TLE.

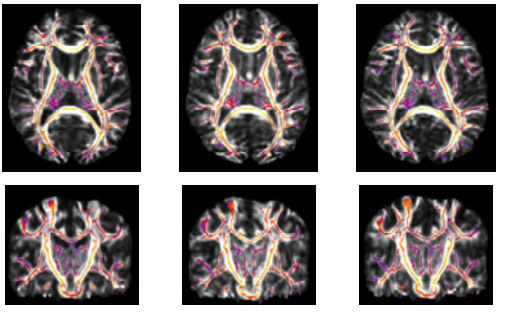


Fig. 4. Examples of the mean FA skeletons in the subject images

3 Results

We applied the FAA method on DTI of patients with partial TLE. TLE was defined in 1985 by the International League Against Epilepsy (ILAE) as a condition characterized by recurrent unprovoked seizures originating from the medial or lateral temporal lobe. Approximately half of patients with epilepsy have partial epilepsy, and partial epilepsy is often of temporal lobe origin. The seizures associated with TLE consist of simple or complex partial seizures. For many patients with complex partial seizures of temporal lobe origin, anterior temporal lobectomy can provide the possibility of seizure-free or nearly seizure-free treatment and reduce dependency on antiepileptic drugs. Accurately analyzing the morphological and white matter tracts in the temporal lobe provide quantitative measures about the status of the disease as well as the outcome of the drug treatment, and also help us to better understand the relationship between seizures and the abnormality of the brain to give a better accuracy of suitable cases for anterior temporal lobectomy.

While EEG is the gold standard for identifying the location of seizure foci during the surgical work-up of patients with intractable TLE, other methods are routinely used to improve diagnostic certainty. DTI can be used to detect the integrity of neural fibers and may be useful when conventional neuroimaging is unable to assist in the corroborating of the side of seizure focus. In experiments, we relate WM tract integrity in TLE with side of seizure focus using the FAA method. 18 patients diagnosed with partial epilepsy (9 left and 9 right were confirmed with EEG) and 12 normal controls were used in this study. For each set of data, we first registered them with the template image and then deformed the FA map of each subject onto the template space using the resultant deformation field. The average FA map was calculated and the FA skeleton was generated thereafter. Finally, the FA asymmetry of each region is calculated using Eq.(4). Statistical analysis was then performed among the three groups. Table 1 illustrates all the regions/tracts contribute significant to the group differences.

Table 1. Statistical testing results

Group comparison	Significant regions contributing group difference (p-value<0.008)
Control v.s. Left TLE	superior, middle, and inferior temporal WM, uncinate fasciculus, superior fronto-occipital fasciculus, middle fronto-orbital WM and lateral fronto-orbital WM.
Control v.s. Right TLE	superior temporal WM, uncinate fasciculus, lateral fronto-orbital WM, fusiform WM.
Left v.s. Right TLE	inferior-frontal WM, superior, middle, and inferior temporal WM.

Overall, we can see that temporal lobe WM contributes the most, and seizures of the temporal lobe origin can have dramatic effects on an individual's personality. This is because temporal lobes are highly associated with language skills. Left temporal abnormality disturbs the recognition of words, while that of right temporal damage can cause an inhibition of talking. This is in agreement with the symptoms like perseverative speech, paranoia, and aggressive rages of TLE [14]. Another important structure is the uncinate fasciculus (UF). Although its function is still unknown, it is believed that the right UF might be associated with autonoetic self awareness, while the left UF may be linked to memory [15]. The third group of WM that can distinguish

TLE is bundle linking frontal lobes. Notice that there are relatively less regions to distinguish right TLE from controls. Overall, there are significant FA asymmetries of the fiber tracts on the temporal lobes as well as associated fibers found using our method, which are in agreement with the symptoms of TLE. With the control and TLE patients' data, this paper determines the regions that contribute significantly to the group differences. We have shown that by using the FAA method, the left and right TLE can be differentiated from controls. With more patient and control data being collected from our hospital, a robust classifier can be designed based on these fiber tracts for automatic classification of TLEs with different site of origin.

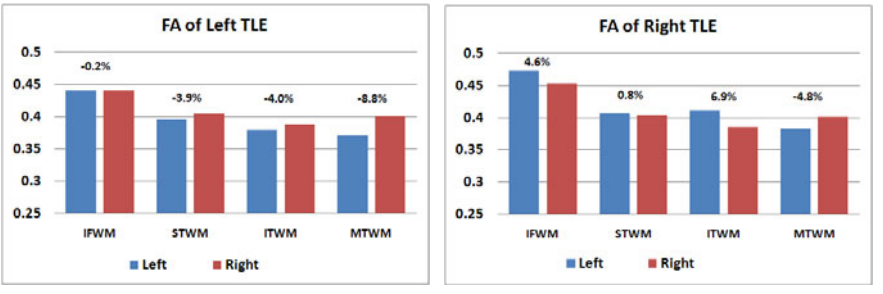


Fig. 5. Average FA values of WM skeleton section pairs in patients with left and right TLE. Blue bars show the values on the left, and red bar show the values on the right ROIs. The percentage shows the difference between them (i.e., (left-right)/right). The ROIs list herein includes IFWM, inferior-frontal WM; STWM, superior temporal WM; ITWM, inferior temporal WM; and MTWM, middle temporal WM.

The FA values of the temporal lobe regions contributing significantly to the group difference between left vs. right TLE are shown in Fig. 5. It can be seen that the FA values of the WM tracks are lower on the side of seizures origin.

In future works, we will correlate the quantitative DTI findings with neuro psychological, EEG, and functional MRI data for increased diagnostic accuracy of seizure lateralization during the pre-surgical evaluation of individuals with TLE.

4 Conclusion

This paper presents an automatic FA asymmetry analysis and applies it to determine the FAA changes associated with the sides of seizure origin of patients with TLE using DTI. First, a simultaneous DTI registration algorithm is utilized to differentiate the control and patient images into the JHU-DTI-MNI atlas, and the FA images are warped accordingly. Then, the TBSS algorithm is employed to quantify the FA on the WM skeletons. Using the atlas, the skeleton is divided into different sections accordingly. The FAA values, i.e., the relative differences of each WM skeleton section pair, are calculated for statistical analysis. In experiments, we have utilized the FAA to identify regions that will significantly contribute to the group differences between the control and the left/right TLE, as well as between left and right TLE groups. We have concluded based on the results indicated that FAA correlates with

the location of seizure origin, and those of certain regions are significantly different between normal controls and left or right TLE. The quantitative results can be useful for pre-surgical evaluation of TLE patients and for better understanding the relationship between fiber tracts with the site of origin of TLE, EEG tests, the syndromes and neural psychological responses.

References

1. Wakana, S., Caprihan, A., Panzenboeck, M.M., Fallon, J.H., Perry, M., Gollub, R.L., Hua, K., Zhang, J., Jiang, H., Dubey, P., Blitz, A., van Zijl, P., Mori, S.: Reproducibility of Quantitative Tractography Methods Applied to Cerebral White Matter. NeuroImage 36, 630–644 (2007)
2. O'Donnell, L., Westin, C.F.: White Matter Tract Clustering and Correspondence in Populations. Med. Image Comput. Assist. Interv. 8, 140–147 (2005)
3. Brun, A., Knutsson, H., Park, H.J., Shenton, M.E., Westin, C.F.: Clustering Fiber Traces Using Normalized Cuts. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3216, pp. 368–375. Springer, Heidelberg (2004)
4. Maddah, M., Grimson, W.E., Warfield, S.K., Wells, W.M.: A Unified Framework for Clustering and Quantitative Analysis of White Matter Fiber Tracts. Medical Image Analysis 12, 191–202 (2008)
5. Ding, Z., Gore, J.C., Anderson, A.W.: Classification and Quantification of Neuronal Fiber Pathways Using Diffusion Tensor MRI. Magn. Reson. Med. 49, 716–721 (2003)
6. Mori, S., Oishi, K., Jiang, H., Jiang, L., Li, X., Akhter, K., Hua, K., Faria, A.V., Mahmood, A., Woods, R., Toga, A.W., Pike, G.B., Neto, P.R., Evans, A., Zhang, J., Huang, H., Miller, M.I., van Zijl, P., Mazziotta, J.: Stereotaxic White Matter Atlas Based on Diffusion Tensor Imaging in an Icbm Template. NeuroImage 40, 570–582 (2008)
7. Smith, S.M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T.E., Mackay, C.E., Watkins, K.E., Ciccarelli, O., Cader, M.Z., Matthews, P.M., Behrens, T.E.: Tract-Based Spatial Statistics: Voxelwise Analysis of Multi-Subject Diffusion Data. NeuroImage 31, 1487–1505 (2006)
8. Yeh, P.-H., Simpson, K., Durazzo, T.C., Gazdzinski, S., Meyerhoff, D.: Tract-Based Spatial Statistics (Tbss) of Diffusion Tensor Imaging Data in Alcohol Dependence: Abnormalities of the Motivational Neurocircuitry. Psychiatry Research: Neuroimaging 173, 22–30 (2009)
9. Xue, Z., Li, H., Guo, L., Wong, S.T.: A Local Fast Marching-Based Diffusion Tensor Image Registration Algorithm by Simultaneously Considering Spatial Deformation and Tensor Orientation. NeuroImage 52, 119–130 (2010)
10. Zhang, H., Yushkevich, P.A., Alexander, D.C., Gee, J.C.: Deformable Registration of Diffusion Tensor MR Images with Explicit Orientation Optimization. Medical Image Analysis 10, 764–785 (2006)
11. Yeo, B.T., Vercauteren, T., Fillard, P., Pennec, X., Golland, P., Ayache, N., Clatz, O.: Dti Registration with Finite-Strain Differential. In: ISBI, pp. 700–703 (2008)
12. Sethian, J.A.: Level Set Methods and Fast Marching Methods. Cambridge University Press, Cambridge (1999)
13. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J.: Nonrigid Registration Using Free-Form Deformations: Application to Breast MR Images. IEEE Transactions on Medical Imaging 18, 712–721 (1999)
14. Trebuchon-Da Fonseca, A., et al.: Brain regions underlying word finding difficulties in temporal lobe epilepsy. Brain 132, 2772–2784 (2009)
15. Diehl, B., et al.: Abnormalities in diffusion tensor imaging of the uncinate fasciculus relate to reduced memory in temporal lobe epilepsy. Epilepsia 49, 1409–1418 (2008)

Hierachical Spherical Harmonics Based Deformable HARDI Registration

Pew-Thian Yap[1], Yasheng Chen[1], Hongyu An[1], John H. Gilmore[2], Weili Lin[1], and Dinggang Shen[1]

[1] BRIC, Department of Radiology and
[2] Department of Pyschiatry
University of North Carolina at Chapel Hill, NC

Abstract. In contrast to the more common Diffusion Tensor Imaging (DTI), High Angular Resolution Diffusion Imaging (HARDI) allows superior delineation of angular microstructures of brain white matter, and makes possible multiple-fiber modeling of each voxel for better characterization of brain connectivity. However, in the context of image registration, the question of how much information is needed for satisfactory alignment remains unanswered. Low order representation of the diffusivity information is generally more robust than the higher order representation, but the latter gives more information for correct fiber tract alignment. However, higher order representation, when naïvely utilized, might not necessarily be conducive to improving registration accuracy since similar structures with significant orientation differences prior to proper alignment might be mistakenly taken as non-matching structures. We propose in this paper a hierarchical spherical harmonics based registration algorithm which utilizes the wealth of information provided by HARDI in a more principled means. The image volumes are first registered using robust, relatively direction invariant features derived from the diffusion-attenuation profile, and their alignment is then refined using spherical harmonic (SH) representation of gradually increasing order. This progression of SH representation from non-directional, single-directional to multi-directional representation provides a systematic means of extracting directional information from the HARDI data. Experimental results show a significant increase in registration accuracy over a state-of-the-art DTI registration algorithm.

1 Introduction

The shortcoming of Diffusion Tensor Imaging (DTI) in resolving intra-voxel multiple fiber crossings has prompted great interest in developing more sophisticated models. Notably, Tuch et al. [1,2] introduced the High Angular Resolution Diffusion Imaging (HARDI) method, suggesting that the apparent diffusion coefficients could be evaluated along many different directions without fitting a global function to the data. The outcome is a diffusivity profile consisting of an angular distribution of apparent diffusivities, which allows a more complex representation of each voxel, and makes capturing multi-fiber information possible. This,

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 228–236, 2010.

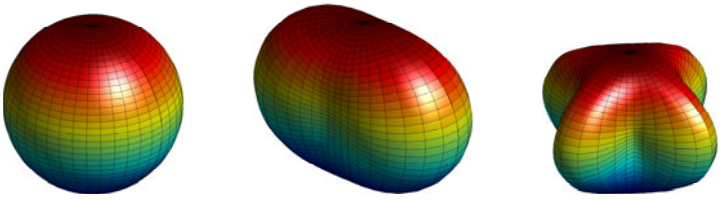


Fig. 1. (a) Zeroth order, (b) second order, and (c) fourth order SHs for representing isotropic, single-fiber and dual-fiber diffusivity profiles

however, presents new problems in the context of image registration. In particular, the question of how much information is actually needed for achieving satisfactory registration is still unanswered.

In view of the exciting insights into the brain HARDI can offer, there has been a recent flourish of HARDI registration algorithms. Fourth order tensors [3], along with a distance called the Hellinger distance, were employed by Barmpoutis et al. [4] for registration of human hippocampi. Geng et al. [5] proposed a method which utilizes a spherical harmonic (SH) representation of the orientation distribution function (ODFs). Reorientation was performed directly on the SH coefficients, in a manner similar to the Finite Strain (FS) tensor reorientation technique proposed by Alexander et al. [6], by extracting rotation matrices from local Jacobians. Cheng et al. [7] took a multi-compartmental approach which was based on Gaussian mixtures. Reorientation was performed on the individual Gaussian components, each representing a major fiber direction, using an extension of the Preservation of Principal Directions (PPD) method [7]. Hong et al. performed registration on the $b = 0$ images and applied the estimated deformation field on the diffusion-weighted images with retransformation — taking into account rotation, scaling, and shearing effects of the spatial transformation — of the fiber orientation distribution. Bloy and Verma [8] computed the SH band energies as features for use with a Demons-based multi-channel registration algorithm for alignment of HARDI data.

For effective utilization of the wealth of information afforded by HARDI, we propose in this paper a hierarchical SHs based HARDI registration algorithm. SHs [9] of different orders have been shown to be capable of representing different diffusivity patterns [10]. Specifically, zeroth, second and fourth order SHs (see Fig. 1) have basis functions with shapes which are sphere-like, cigar-like and cross-like — ideal for representing voxels with isotropic, single-fiber, and dual-fiber diffusivity profiles, respectively. This property presents a natural way of hierarchically extracting information from the diffusivity profile. The key of our approach involves representing the diffusion-attenuation profile with increasing complexity to help progressively refine the registration. This effectively allows lower order, relatively orientation invariant, but more robust representations to guide the initial registration and higher order, directional, more precise representations to refine the alignment. This approach puts high and low order information in their

proper contexts and helps avoid mismatching of diffusion directions arising from large structural orientation differences. Registration is achieved by employing a forward-backward-consistent soft-correspondence matching scheme.

2 Diffusion-Attenuation Descriptors

2.1 Spherical Harmonic Representation

SHs, denoted as Y_l^m, with l denoting the order and m the phase factor, are a basis for complex functions on the unit sphere. Explicitly, they are given as:

$$Y_l^m(\theta,\phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}} P_l^m(\cos(\theta)) e^{im\phi} \tag{1}$$

where (θ,ϕ) obeys the physics convention $(\theta \in [0,\pi]), \phi \in [0,2\pi])$ and P_l^m is an associated Legendre polynomial. Since diffusion signals are real and antipodal symmetric, it is sufficient to utilize a real basis function set of even orders, i.e., for $l = 0, 2, 4, \ldots, L$ and $m = -L, \ldots, 0, \ldots, L$:

$$Y_l^m = \begin{cases} \sqrt{2} \cdot \mathrm{Re}\left(Y_l^m\right), & -L \le m < 0 \\ Y_l^0, & m = 0 \\ \sqrt{2} \cdot \mathrm{Im}\left(Y_l^m\right), & 0 < m \le L \end{cases} \tag{2}$$

where $\mathrm{Re}\left(Y_l^m\right)$ and $\mathrm{Im}\left(Y_l^m\right)$ represent the real and imaginary parts of Y_l^m, respectively.

We denote $S(\mathbf{g})$ as the diffusion-attenuation signals at a finite set of points on a sphere, where $\mathbf{g}$ is the diffusion encoding direction in a pulsed-gradient spin-echo experiment. As a spherical function, it can be approximated as a linear combination of a set of SH basis functions:

$$S(\mathbf{g}) = \sum_{l=0}^{L} \sum_{m=-l}^{l} y_l^m Y_l^m(\mathbf{g}). \tag{3}$$

Order L determines the complexity of the representation. A higher value for L will bring the representation closer to the original signal and will hence extract more directional information [10]. The SH coefficients y_l^m are determined using a Laplace-Beltrami regularized least-squares estimator, as described in [11].

We note here that the diffusion-attenuation profile $S(\mathbf{g})$ has a shape which is different compared to the diffusivity profile used in [9,10] (see [11] for examples). For example, a voxel encoding a single fiber direction will have an oblate shape in contrary to the typical prolate shape of the diffusivity profile. However, one can still sufficiently represent the single-fiber-direction voxel by SHs up to the second order. The same goes for voxels with fiber crossings. SH representation of $S(\mathbf{g})$ was in fact used in [12,11] for computing the ODFs. Funk-Radon transform, working in tandem with Funk-Hecke theorem [11], allows us to obtain the ODF

by using scaled versions of the SH coefficients [11], i.e., $2\pi P_l(0) y_l^m$, where $P_l(x)$ is the Legendre polynomial of degree l. That is, the ODF given by $S(\mathbf{g})$ is:

$$\mathcal{G}[S](\mathbf{g}) = \sum_{l=0}^{L} \sum_{m=-l}^{l} 2\pi P_l(0) y_l^m Y_l^m(\mathbf{g}). \tag{4}$$

2.2 Statistical Descriptors and Edge Maps

We incorporate, in addition to SHs, the following statistical descriptors of the diffusion-attention profiles in the registration algorithm. The diffusion mean is computed over the unit sphere Ω:

$$\mu = \langle S(\mathbf{g}) \rangle = \frac{1}{4\pi} \int_{\mathbf{g} \in \Omega} S(\mathbf{g}) d\mathbf{g} = \frac{y_0^0}{\sqrt{4\pi}}, \tag{5}$$

characterizing the average diffusion magnitude. The diffusion deviation, which measures the deviation of the profile from its isotropic component, is defined as:

$$\rho^2 = \frac{\left\langle \left| S(\mathbf{g}) - \langle S(\mathbf{g}) \rangle \right|^2 \right\rangle}{\left\langle \left| S(\mathbf{g}) \right|^2 \right\rangle} = \frac{\sum_{l=1}^{L} \sum_{m=-l}^{l} \left| y_l^m \right|^2}{\sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left| y_l^m \right|^2}. \tag{6}$$

For better characterization of structural shapes, regional statistical descriptors consisting of regional diffusion mean:

$$\mu_{\mathcal{N}} = \langle \mu(\mathbf{z}) \rangle = \frac{1}{|\mathcal{N}|} \sum_{z \in \mathcal{N}} \mu(\mathbf{z}) \tag{7}$$

and regional diffusion deviation:

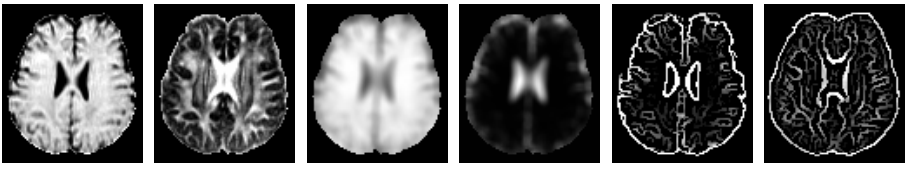
$$\rho_{\mathcal{N}}^2 = \langle \sigma(\mathbf{z}) \rangle = \frac{\sum_{\mathbf{z} \in \mathcal{N}} \left| \mu(\mathbf{z}) - \mu_{\mathcal{N}} \right|^2}{\sum_{\mathbf{z} \in \mathcal{N}} \left| \mu(\mathbf{z}) \right|^2} \tag{8}$$

are computed for the neighborhood $\mathcal{N}$ of each voxel. Although the above measures can be computed via the SH coefficients, considerable time can be saved by direct computation using $S(\mathbf{q})$. In addition to these descriptors, we applied a Canny edge detector on the μ and ρ maps to obtain edge information (denoted as $\mathcal{H}_\mu$, $\mathcal{H}_\rho$) for guiding the alignment of tissue boundaries. Example statistical and edge maps are shown in Fig. 2.

2.3 Feature Vector and Similarity Measure

For each voxel at location $\mathbf{x}$, the SH coefficients and diffusion-attenuation statistical descriptors are grouped into a vector $\mathbf{a}(\mathbf{x}) = [\mathbf{a}_1(\mathbf{x}), \mathbf{a}_2(\mathbf{x})]$, with:

$$\mathbf{a}_1(\mathbf{x}) = \Big[\mu(\mathbf{x}), \rho(\mathbf{x}), \mu_{\mathcal{N}}(\mathbf{x}), \rho_{\mathcal{N}}(\mathbf{x}), \mathcal{H}_\mu(\mathbf{x}), \mathcal{H}_\rho(\mathbf{x}) \Big], \ \mathbf{a}_2(\mathbf{x}) = \Big[\{ w_l y_l^m(\mathbf{x}) \} \Big] \tag{9}$$



(a) (b) (c) (d) (e) (f)

Fig. 2. Statistical and edge maps: (a) Mean, μ, (b) Deviation, ρ, (c) Regional mean, $\mu_{\mathcal{N}}$, (d) Regional deviation, $\rho_{\mathcal{N}}$, (e) μ edge map, $\mathcal{H}_\mu$ and (f) ρ edge map, $\mathcal{H}_\rho$

where $\{w_l y_l^m(\mathbf{x})\}$ is the set of weighted SH cofficients computed up to order L. The weight of each order-band l is given by w_l ($w_0 = 1 \geq w_{l-2} \geq w_l \geq 0$). For our implemetation, we set $w_2 = u(\alpha_2 \tau)$ and $w_4 = u(\alpha_4 \tau)$, where $u(\cdot)$ is a step function, for allowing the matching of single-fiber and dual-fiber information after α_2, $\alpha_4 \in [0,1]$ ($\alpha_2 < \alpha_4$) fractions of the total number of matching iterations, τ, respectively. This is so that progressively higher order representations can be employed to refine the registration. Normalizing the elements of $\mathbf{a}_1(\mathbf{x})$ to have a range of $[0,1]$, the similarity measure can be defined as:

$$\eta(\mathbf{x}_\mathrm{T}, \mathbf{x}_\mathrm{S}) = \left\{ \Theta\left[\mathbf{a}_1(\mathbf{x}_\mathrm{T}) - \mathbf{a}_1(\mathbf{x}_\mathrm{S})\right] \right\} \times \left\{ \frac{\mathbf{a}_2(\mathbf{x}_\mathrm{T}) \cdot \mathbf{a}_2(\mathbf{x}_\mathrm{S})}{||\mathbf{a}_2(\mathbf{x}_\mathrm{T})|| \times ||\mathbf{a}_2(\mathbf{x}_\mathrm{S})||} \right\} \tag{10}$$

where $\Theta[\mathbf{a}] = \Pi_k(1 - |a_k|)$, and $\mathbf{x}_\mathrm{T}$ and $\mathbf{x}_\mathrm{S}$ are specific voxels in the template image and the subject image, respectively. For robustness, we employ instead regional similarity measures:

$$\begin{aligned} \eta_{\mathcal{M}}(\mathbf{x}_\mathrm{T}, \mathbf{x}_\mathrm{S}) &= \frac{1}{|\mathcal{M}(\mathbf{x}_\mathrm{T})|} \sum_{\mathbf{z} \in \mathcal{M}(\mathbf{x}_\mathrm{T})} \eta(\mathbf{z}, \mathbf{x}_\mathrm{S} + f(\mathbf{z}) - f(\mathbf{x}_\mathrm{T})) \\ \eta_{\mathcal{M}}(\mathbf{x}_\mathrm{S}, \mathbf{x}_\mathrm{T}) &= \frac{1}{|\mathcal{M}(\mathbf{x}_\mathrm{S})|} \sum_{\mathbf{z} \in \mathcal{M}(\mathbf{x}_\mathrm{S})} \eta(\mathbf{x}_\mathrm{T} + f^{-1}(\mathbf{z}) - f^{-1}(\mathbf{x}_\mathrm{S}), \mathbf{z}) \end{aligned} \tag{11}$$

which compare the similarity of the feature vectors in the neighborhood $\mathcal{M}$ surrounding $\mathbf{x}_\mathrm{T}$ in the template image with that of $\mathbf{x}_\mathrm{S}$ in the subject image with consideration of the transformation f. Note that similariy measures in both directions are defined so that they can be employed for forward-backward consistent matching, described in the next section. The SH coefficients are re-estimated in each iteration of the registration to take into account the effect of the spatial transformation has on the diffusion-attenuation profile.

3 HARDI Registration

Registration of the HARD images can be achieved by minimizing the following cost function:

$$C(\mathbf{P}_\mathrm{T}, \mathbf{P}_\mathrm{S}, f) = \underbrace{C(\mathbf{P}_\mathrm{T}, f) + C(\mathbf{P}_\mathrm{S}, f)}_{\text{Matching}} + \underbrace{C(\mathbf{P}_\mathrm{T}) + C(\mathbf{P}_\mathrm{S})}_{\text{Soft Correspondence}} + \underbrace{C(f)}_{\text{Smoothness}} . \tag{12}$$

Matching: Voxel at $\mathbf{x}_\mathrm{T}$ and voxel at $\mathbf{x}_\mathrm{S}$ are deemed as a matching pair if they are close spatially and if their feature vectors show high similarity. Naturally, a voxel pair $\mathbf{x}_\mathrm{T}$ and $\mathbf{x}_\mathrm{S}$ satisfying these conditions will be given a higher probability value $p(\mathbf{x}_\mathrm{T}, \mathbf{x}_\mathrm{S})$ in the cost function. We also enforce a symmetric matching mechanism [13] which avoids bias towards the template or the subject. The cost functions are defined as:

$$\begin{aligned} C(\mathbf{P}_\mathrm{T}, f) &= \sum_{\mathbf{x}_\mathrm{T}\in\mathcal{V}_\mathrm{T}, \mathbf{x}_\mathrm{S}\in\mathcal{V}_\mathrm{S}} p_\mathrm{T}(\mathbf{x}_\mathrm{T}, \mathbf{x}_\mathrm{S})\Big\{||f(\mathbf{x}_\mathrm{T}) - \mathbf{x}_\mathrm{S}||^2 - \log\left[\eta_{\mathcal{M}}(\mathbf{x}_\mathrm{T}, \mathbf{x}_\mathrm{S})\right]\Big\} \\ C(\mathbf{P}_\mathrm{S}, f) &= \sum_{\mathbf{x}_\mathrm{T}\in\mathcal{V}_\mathrm{T}, \mathbf{x}_\mathrm{S}\in\mathcal{V}_\mathrm{S}} p_\mathrm{S}(\mathbf{x}_\mathrm{T}, \mathbf{x}_\mathrm{S})\Big\{||\mathbf{x}_\mathrm{T} - f^{-1}(\mathbf{x}_\mathrm{S})||^2 - \log\left[\eta_{\mathcal{M}}(\mathbf{x}_\mathrm{S}, \mathbf{x}_\mathrm{T})\right]\Big\} \end{aligned} \tag{13}$$

where $\mathbf{P}_\mathrm{T} = \{p_\mathrm{T}(\mathbf{x}_\mathrm{T}, \mathbf{x}_\mathrm{S})\}$ and $\mathbf{P}_\mathrm{S} = \{p_\mathrm{S}(\mathbf{x}_\mathrm{T}, \mathbf{x}_\mathrm{S})\}$. $\mathcal{V}_\mathrm{T}$ and $\mathcal{V}_\mathrm{S}$ represent the template and subject brain domains, respectively.

Soft Correspondence: Soft correspondence are permitted in the initial stages of the registration to allow robust matching based on multiple candidate points. Towards the end of the registration, more exact one-to-one correspondence is enforced. This is realized by energy terms:

$$\begin{aligned} C(\mathbf{P}_\mathrm{T}) &= \gamma \sum_{\mathbf{x}_\mathrm{T}\in\mathcal{V}_\mathrm{T}, \mathbf{x}_\mathrm{S}\in\mathcal{V}_\mathrm{S}} p_\mathrm{T}(\mathbf{x}_\mathrm{T}, \mathbf{x}_\mathrm{S}) \log(p_\mathrm{T}(\mathbf{x}_\mathrm{T}, \mathbf{x}_\mathrm{S})), \\ C(\mathbf{P}_\mathrm{S}) &= \gamma \sum_{\mathbf{x}_\mathrm{T}\in\mathcal{V}_\mathrm{T}, \mathbf{x}_\mathrm{S}\in\mathcal{V}_\mathrm{S}} p_\mathrm{S}(\mathbf{x}_\mathrm{T}, \mathbf{x}_\mathrm{S}) \log(p_\mathrm{S}(\mathbf{x}_\mathrm{T}, \mathbf{x}_\mathrm{S})). \end{aligned} \tag{14}$$

where $\mathcal{V}_\mathrm{T}$ and $\mathcal{V}_\mathrm{S}$ represent the template and subject brain domains, respectively. Parameters γ controls the degree of fuzziness of the matching. It has initially high values, encouraging fuzzy matching, and later progressively lower values, which enforce exact matching.

Transformation Regularization: Mapping f is required to be smooth for preserving a biologically sensible topology. This is enforced by energy term: $E(f) = \beta||\mathcal{L}f||^2$. $\mathcal{L}$ is an operator which aids in measuring the bending energy. β is a weighting factor which is decreased throughout the registration to allow f to model deformation of increasing complexity.

Optimization: The cost function (12) can be minimized by alternating between correspondence matching and dense transformation estimation [14,15]. We first fix f and solve for $\mathbf{P}_\mathrm{T}$ and $\mathbf{P}_\mathrm{S}$ by letting $\partial C(\mathbf{P}_\mathrm{T}, \mathbf{P}_\mathrm{S}, f)/\partial p_\mathrm{T}(\mathbf{x}_\mathrm{T}, \mathbf{x}_\mathrm{S}) = 0$ and $\partial C(\mathbf{P}_\mathrm{T}, \mathbf{P}_\mathrm{S}, f)/\partial p_\mathrm{S}(\mathbf{x}_\mathrm{T}, \mathbf{x}_\mathrm{S}) = 0$. We then fix $\mathbf{P}_\mathrm{T}$ and $\mathbf{P}_\mathrm{S}$, and solve for f using thin-plate splines (TPS) [16,17].

Retransformation of Spherical Harmonics: Barmpoutis et al. [4] noted the limitations of reorientation of diffusivity functions and proposed instead to employ a full affine *retransformation*. In our case, although a SH reorientation strategy was proposed by Geng et al. [5], we have opted accuracy over speed or convenience by re-estimating the SH coefficients in each iteration of registration. That is, for a local tranform $\mathbf{F}$, we tilt the gradient directions by $\mathbf{g}' = \mathbf{Fg}/||\mathbf{Fg}||$ (see Fig. 3), and compute a new set of SH coefficients $\{{y'}_l^m\}$.

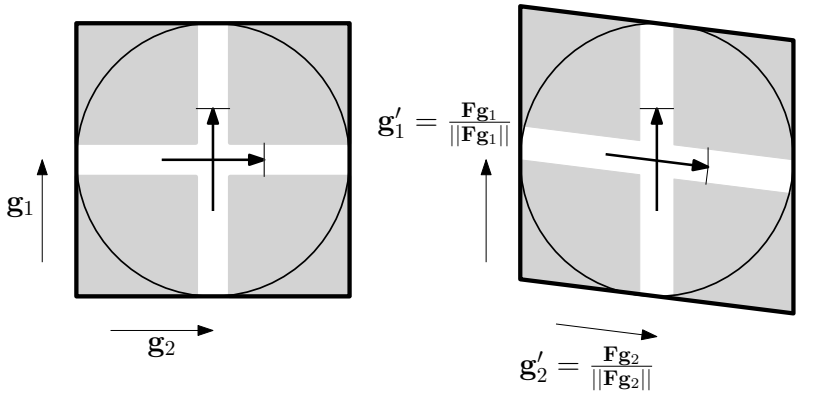


Fig. 3. Retransformation of diffusion-attenuation profile. Each diffusion direction $\mathbf{g}$ is transformed by the local transformation matrix $\mathbf{F}$ to become $\mathbf{g}'_i = \mathbf{F}\mathbf{g}_i/||\mathbf{F}\mathbf{g}_i||$.

4 Experimental Results

Six adult subjects were scanned using an EPI sequence with diffusion gradients applied in 120 non-collinear directions. 80 contiguous slices with slice thickness of 2mm covered a field of view (FOV) of $256 \times 256\text{mm}^2$ with an isotropic voxel size of 2mm. Out of the 6 HARD images, one was selected as the template onto which the rest were registered. To demonstrate the effectiveness of the proposed method over DTI-based registration, we employed a state-of-the-art DTI registration algorithm [15,14] for comparison. We reconstructed the diffusion tensors (DTs) of the diffusion-weighted images and register the DT images using the above-mentioned algorithm. The estimated deformation fields were then used to warp and retransform the respective HARD images. We have also compared our results with FLIRT [18] applied on the deviation maps (see Fig. 2(b)) for affine registration. The orientation consistency (OC) was assessed using the voxel-wise scalar product of the principal orientation direction, $\mathbf{g}_{\text{PD}}$, of each voxel in an aligned image with respect to the template:

$$\text{OC} = \sum_{\mathbf{x} \in \mathcal{V}_{\text{T}}} |\mathbf{g}_{\text{PD}}^{\text{subject}}(\mathbf{x}) \cdot \mathbf{g}_{\text{PD}}^{\text{template}}(\mathbf{x})|. \tag{15}$$

The absolute value was taken since diffusion directions are antipodal symmetric. The principal directions were estimated by a 1082 point even sampling of the orientation distribution functions (ODFs) using Camino [19]. Comparison was made only for voxel showing significant anisotropy (voxels with deviation value $\rho > 0.1$). Compared with DTI-based registration, we found an overall improvement in white-matter orientational consistency of 3.02% ($p < 0.05$). Compared with affine registration, an improvement of 7.39% ($p < 0.001$) was achieved. Results for each individual subject, shown in Fig. 4, indicate that the proposed method gives consistent improvement across subjects.

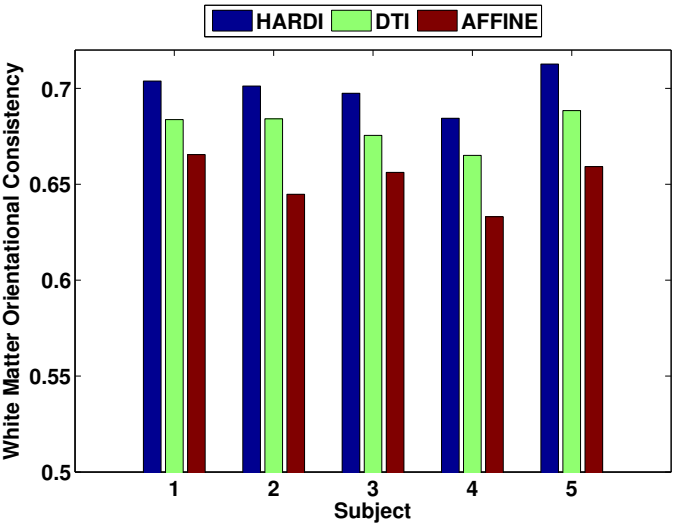


Fig. 4. White matter orientational consistency for individual subjects

5 Conclusion

We have proposed a HARDI registration algorithm which, in a principled manner, extracts orientation information from the diffusion-attenuation profile. The useful property of spherical harmonics in representing different levels of orientation complexity of the diffusion pattern is utilized to progressively provide the registration process with increasing levels of information for alignment refinement. Experimental results indicate better registration accuracy than DTI-based registration algorithm, further validating the fact that the information given by HARDI can be utilized to improve structural alignment.

Acknowledgment

This work was supported in part by grants EB006733, EB008760, EB008374, EB009634, and MH088520.

References

1. Tuch, D.S., Weisskoff, R.M., Belliveau, J.W., Wedeen, V.J.: High angular resolution diffusion imaging of the human brain. In: ISMRM 1999 (1999)
2. Tuch, D.: Q-ball imaging. Magnetic Resonance in Medicine 52(6), 1358–1372 (2004)
3. Barmpoutis, A., Hwang, M.S., Howland, D., Forder, J.R., Vemuri, B.C.: Regularized positive-definite fourth order tensor field estimation from DW-MRI. NeuroImage 45, 153–162 (2009)

4. Barmpoutis, A., Vemuri, B.C., Forder, J.R.: Registration of high angular resolution diffusion MRI images using 4th order tensors. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 908–915. Springer, Heidelberg (2007)
5. Geng, X., Ross, J.T., Zhan, W., Gu, H., Chao, Y.P., Lin, C.P., Christensen, G.E., Schuff, N., Yang, Y.: Diffusion MRI registration using orientation distribution functions. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 627–637. Springer, Heidelberg (2009)
6. Alexander, D.C., Pierpaoli, C., Basser, P.J., Gee, J.C.: Spatial transformations of diffusion tensor magnetic resonance images. IEEE Transactions on Medical Imaging 20(11), 1131–1139 (2001)
7. Cheng, G., Vemuri, B.C., Carney, P.R., Mareci, T.H.: Non-rigid registration of high angular resolution diffusion images represented by Gaussian mixture fields. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 190–197. Springer, Heidelberg (2009)
8. Bloy, L., Verma, R.: Demons registration of high angular resolution diffusion images. In: Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI 2010), pp. 1013–1016 (2010)
9. Alexander, D., Barker, G., Arridge, S.: Detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data. Magnetic Resonance in Medicine 48, 331–340 (2002)
10. Frank, L.R.: Characterization of anisotropy in high angular resolution diffusion-weighted MRI. Magnetic Resonance in Medicine 47, 1083–1099 (2002)
11. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Regularized, fast, and robust analytical q-ball imaging. Magnetic Resonance in Medicine 58, 497–510 (2007)
12. Hess, C.P., Mukherjee, P., Han, E.T., Xu, D., Vigneron, D.B.: Q-ball reconstruction of multimodel fiber orientations using the spherical harmonic basis. Magnetic Resonance in Medicine 56, 104–117 (2006)
13. Christensen, G.E.: Consistent linear-elastic transformations for image matching. In: Information Processing in Medical Imaging, pp. 224–237 (1999)
14. Yap, P.T., Wu, G., Zhu, H., Lin, W., Shen, D.: Fast Tensor Image Morphing for Elastic Registration. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 721–729. Springer, Heidelberg (2009)
15. Yap, P.T., Wu, G., Zhu, H., Lin, W., Shen, D.: F-TIMER: Fast Tensor Image Morphing for Elastic Registration. IEEE Transactions on Medical Imaging 29, 1192–1203 (2010)
16. Bookstein, F.L.: Principal warps: Thin-plate splines and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence 11(6), 567–585 (1989)
17. Chui, H., Rangarajan, A.: A new point matching algorithm for non-rigid registration. Computer Vision and Image Understanding 89(2-3), 114–141 (2003)
18. Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., Luca, M.D., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., Stefano, N.D., Brady, J.M., Matthews, P.M.: Advances in functional and structural MR image analysis and implementation as fsl. NeuroImage (23), S208–S219 (2004)
19. Cook, P.A., Bai, Y., Nedjati-Gilani, S., Seunarine, K.K., Hall, M.G., Parker, G.J., Alexander, D.C.: Camino: Open-source diffusion-MRI reconstruction and processing. In: 14th Scientific Meeting of the International Society for Magnetic Resonance in Medicine, vol. 2759 (2006)

Marker-Free Registration for Electromagnetic Navigation Bronchoscopy under Respiratory Motion

Marco Feuerstein[1,2], Takamasa Sugiura[2], Daisuke Deguchi[2], Tobias Reichl[1,2], Takayuki Kitasaka[3], and Kensaku Mori[2,4]

[1] Computer Aided Medical Procedures (CAMP), Technische Universität München
[2] Graduate School of Information Science, Nagoya University
[3] Faculty of Information Science, Aichi Institute of Technology
[4] Information and Communications Headquarters, Nagoya University

Abstract. Electromagnetic navigation bronchoscopy requires the accurate registration of a preinterventional computed tomography (CT) image to the coordinate system of the electromagnetic tracking system. Current state-of-the-art registration methods are manual or do not explicitly take patient's respiratory motion and exact airway shape into account, leading to relatively low accuracy. This paper presents an automated registration method addressing these issues. Electromagnetic tracking data recorded during bronchoscopic examination is matched to the airways by an optimizer utilizing the Euclidean distance map to the centerline of the airways for automated registration. Using a cutaneous sensor on the chest of the patient allows us to approximate respiratory motion by a linear deformation model and adopt the registration result in real time to the current respiratory phase. A thorough *in silico* evaluation on real patient data including CT images taken in 10 respiratory phases shows the significant registration error decrease of our method compared to the current state of the art, reducing the error from 3.5 mm to 2.8 mm.

1 Introduction

Bronchoscopy is a useful tool in the diagnosis and treatment of lung and bronchus cancer, for example to perform transbronchial biopsies of suspicious lymph nodes or pulmonary nodules. However, due to the complexity of the airways and the limited view during bronchoscopy, it is still difficult for a bronchoscopist to advance the bronchoscope and the biopsy needle to a peripheral target without the aid of fluoroscopy. To overcome this limitation, navigation systems have been proposed that localize and visualize the bronchoscope camera and biopsy needle in relation to a preinterventional computed tomography (CT) image and predefined targets and paths within the image [1,2]. Accurate navigation is essential to guide camera and needle to small targets of only a few millimeters or centimeters.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 237–246, 2010.

Such a navigation system requires continuous tracking of the bronchoscope, for which various techniques have been proposed. Image based techniques try to register virtual camera images generated from preinterventional computed tomography data to the real images acquired by the bronchoscope camera [3,4], which can be time-consuming and fail to continuously track the camera motion. Continuous and real-time electromagnetic tracking (EMT) of a small sensor coil attached to the bronchoscope tip can resolve these issues. Promising initial clinical results have been achieved using such an electromagnetic navigation bronchoscopy system [1].

Electromagnetic navigation bronchoscopy requires the registration between the coordinate system of the preinterventional CT image and that of EMT. In a commercially available system, this is performed by identifying about 5 to 9 landmarks in CT image coordinates by mouse clicks and in EMT coordinates by measuring their corresponding points with the sensor [1]. This is an unnatural and time-consuming process of several minutes and does not consider deformations of the airways caused by patient motion, mainly due to respiratory motion.

To deal with patient motion, Gergel et al. propose to apply particle filtering to each camera position acquired via EMT and project it to a previously segmented centerline of the airways [5], while Soper et al. combine electromagnetic tracking, image based tracking, Kalman-filtering, and a respiratory motion compensation method utilizing a surrogate sensor [6]. However, these techniques still require a manual marker-based registration between the CT and the EMT coordinate system. Automated registration methods were therefore proposed, which collect EMT data during bronchoscope movement inside the airways and match this data with airways previously extracted in CT images to obtain a global rigid transformation [7,8]. Even if virtual breathing motion is added for evaluation in a static phantom [9], the resulting transformation matrix does not incorporate dynamic deformations caused by e.g. patient breathing, making the registration inaccurate. Furthermore, some methods [7,9] assume all branches of the airway tree utilized during matching to be straight, which is rarely true for most airways.

In this paper we address the issues of previously proposed automated registration approaches [7,8,9]. We present a marker-free method that incorporates respiratory motion information and a better representation of the airway shape into the registration process. This significantly increases registration and navigation accuracy, as shown in our *in silico* evaluation.

2 Method

A part of our method follows previous works [7,8,9]. First the major airway branches, their centerline, and their tree structure are automatically extracted from a preinterventional CT image of the patient. Before bronchoscopy, an EMT sensor is inserted into the working channel of the bronchoscope and advanced to its tip or, alternatively, fully integrated into the bronchoscope tip not to obstruct the working channel. During bronchoscopic examination of the major airway branches, EMT data of this sensor is recorded and matched to the extracted

airways. Compared to previous works, our new method presented here refines the matching strategy and takes respiratory motion into account, which significantly improves registration and navigation accuracy.

2.1 Respiratory Phase Detection

To acquire surrogate data for the estimation of respiratory motion we use a second cutaneous EMT sensor attached to the patient's chest [6,10], so for each bronchoscope sensor measurement we can obtain a corresponding cutaneous sensor measurement. The main purpose of this cutaneous sensor is to estimate a linear mapping between its motion and the patient's respiratory motion.

When our matching procedure is started, a principal component analysis is performed on all cutaneous sensor positions obtained to date to compute their principal motion axis. All cutaneous sensor positions are then projected onto this axis and scaled to lie between 0 and 1, giving our surrogate data. Its local minima and maxima approximately correspond to full patient inspiration and expiration or vice versa. As we can be fairly safe to say that also the CT image was taken in either inspiration or expiration breath hold, as this is the standard procedure for chest CT, we now only need to determine, whether the minima of the ground truth data correspond to inspiration or expiration, which is done automatically in the next step.

2.2 Rigid Registration

During rigid registration, we compute the Euclidean transformation ${}^{\mathrm{CT}}\boldsymbol{T}_{\mathrm{EMT}}$ from EMT coordinates to CT coordinates using the bronchoscope sensor poses closest to the approximate respiratory phase the CT was acquired in. As CT data sets used for navigation can come from various hospitals and scanners sometimes without any information about whether their respiratory phase was inspiration or expiration, to estimate the correct phase we simply perform two rigid registrations, one only including bronchoscope sensor poses corresponding to surrogate sensor data between 0 and 0.1 and the other one only including bronchoscope sensor poses corresponding to surrogate data between 0.9 and 1. We then simply select the resulting transformation, which better corresponds to the CT data.

Each rigid registration is executed following the method proposed in [9], which can be seen as an iterative closest point-like approach, where all EMT points gradually converge to the airway tree. However, we can greatly improve registration performance by changing the error term used in [9] during optimization of the transformation matrix. Instead of finding the closest point on the straight line segments of the tree representation of the airways to a bronchoscope sensor position (transformed into CT coordinates) and computing their distance, we minimize the distance to the curved airway centerline. This can be achieved effectively by generating a Euclidean distance map d to the centerline obtained during airway segmentation, which is squared and normalized by division of each voxel by the average radius of the branch closest to the voxel. This gives less weight to thick branches than to thin ones, and corresponds to the fact that

bronchoscope movement naturally deviates much more from the centerline in thick branches than in thin branches.

In detail, our new error term is

$$Err\left({}^{\mathrm{CT}}\boldsymbol{T}_{\mathrm{EMT}}\right) = \sum_{\boldsymbol{p}_k \in \mathcal{P}, s_{\min} \leq s(\boldsymbol{p}_k) \leq s_{\max}} \frac{1}{r_k} \cdot d^2\left({}^{\mathrm{CT}}\boldsymbol{T}_{\mathrm{EMT}} \boldsymbol{p}_k\right), \tag{1}$$

where $\boldsymbol{p}_k$ is the kth of N bronchoscope sensor positions, ${}^{\mathrm{CT}}\boldsymbol{T}_{\mathrm{EMT}}$ transforms $\boldsymbol{p}_k$ from EMT to CT coordinates, $s(\boldsymbol{p}_k)$ gives the surrogate data (between 0 and 1) corresponding to $\boldsymbol{p}_k$, $s_{\min}$ and $s_{\max}$ are set once to 0 and 0.1 and once to 0.9 and 1, respectively, r_k is the average radius of the branch closest to $\boldsymbol{p}_k$, and $d(\boldsymbol{x})$ returns the distance of our Euclidean distance map for a point $\boldsymbol{x}$. To determine the six degrees of freedom of ${}^{\mathrm{CT}}\boldsymbol{T}_{\mathrm{EMT}}$, the error is minimized using the CONDOR algorithm [11].

After performing the optimization twice, we divide each resulting error by the number of sensor positions. The lower normalized error will naturally correspond to the sensor data acquired in the respiratory phase close to the phase the CT was acquired in. Accordingly, in the following we choose the resulting transformation ${}^{\mathrm{CT}}\boldsymbol{T}_{\mathrm{EMT}}$ with lower error and set the respiratory phase p_{CT} of CT acquisition to either 0 (approximately corresponding to the result using surrogate data between 0 and 0.1) or 1 (for surrogate data between 0.9 and 1).

2.3 Respiratory Motion Correction

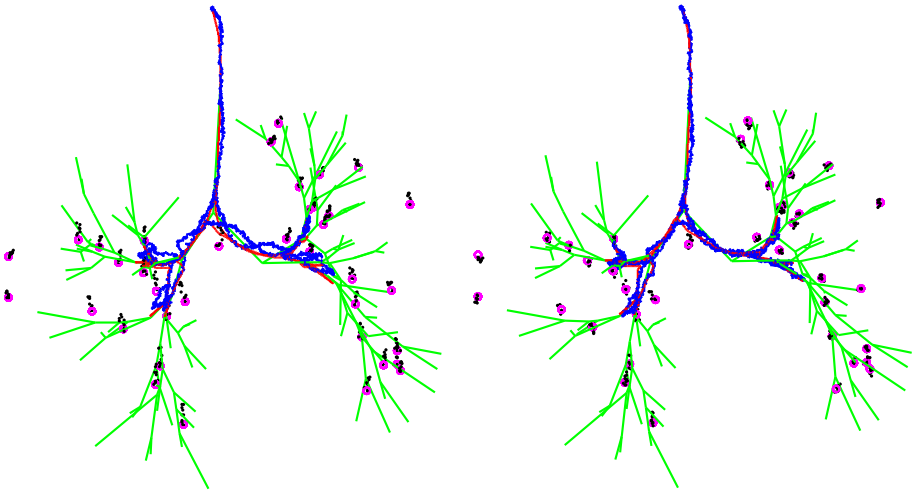
After determination of the best respiratory phase and rigid matching between CT and EMT coordinates, we now apply an additional translation $\boldsymbol{t}_{\mathrm{cor}}$ that is scaled linearly with the respiratory phase measured by the surrogate sensor to correct for breathing motion. Even though this is just a rough approximation of the real deformation that is in fact a spatially varying deformation field, we will see later in our experiments that our simple linear scaling approach can already estimate the main bronchial motion well, as also shown in [6]. It can greatly improve registration accuracy without knowing a dense patient-specific deformation field that is rarely available for navigated bronchoscopy because of its necessity for several CT images of two or more respiratory phases.

In this second optimization step we again utilize CONDOR [11] to find a translation $\boldsymbol{t}_{\mathrm{cor}}$ that minimizes the error

$$Err\left(\boldsymbol{t}_{\mathrm{cor}}\right) = \sum_{\boldsymbol{p}_k \in \mathcal{P}} \frac{1}{r_k} \cdot d^2 \underbrace{\left(\begin{pmatrix} \boldsymbol{I} & \|p_{\mathrm{CT}} - s(\boldsymbol{p}_k)\| \, \boldsymbol{t}_{\mathrm{cor}} \\ \boldsymbol{0}^T & 1 \end{pmatrix} {}^{\mathrm{CT}}\boldsymbol{T}_{\mathrm{EMT}} \boldsymbol{p}_k\right)}_{=:\boldsymbol{p}_{k_{\mathrm{cor}}}}. \tag{2}$$

3 Evaluation

Our institution does not yet permit us to evaluate our experimental electromagnetic navigation bronchoscopy system *in vivo*. Moreover, it is not a trivial task



(a) Simulated data before registration **(b)** After registration

Fig. 1. Results of our *in silico* evaluation, showing airway tree line segments in green, a ground truth bronchoscope path in red, a simulated path (before and after registration) in blue, ground truth landmarks in pink, and their deformed counterparts (before and after registration) in 10 breathing phases in black

to generate ground-truth data and perform a quantitative evaluation of electromagnetic navigation bronchoscopy in the operating room without significantly changing the current clinical work flow. This may be solved by deformable physical phantoms, but they are still far from realistic respiratory motion modeling, and for those the quantification of ground truth motion is still an open question, too. We hence decided to set up a thorough *in silico* evaluation on real patient CT data with exhaustive and accurate breathing information. Therefore we use the POPI model [12], which contains 10 respiratory phases and provides dense deformation fields between all phases as well as landmarks chosen by medical experts and widely distributed inside the lungs and around the airways[1].

We selected the end-inhalation phase of the POPI data to serve as the single CT image that is usually acquired before bronchoscopy. In this CT image we extracted the airways, their centerline, and their tree structure (marked green in Fig. 1a) using our previously proposed method [13], and simulated 10 different bronchoscope paths and their corresponding bronchoscope EMT sensor data. Each bronchoscope path was created using a few manually selected keyframes in every branch close to, but not on its centerline and connecting the frames by Catmull-Rom splines and Slerp quaternion interpolation. It can be adjusted to cover a certain number of airway generations, e.g. 1 referring to the trachea, 2 to the trachea and left and right main bronchi, 3 to the trachea, left and right

[1] The data was obtained from the Léon Bérard Cancer Center & CREATIS lab, Lyon, France; `http://www.creatis.insa-lyon.fr/rio/popi-model`

main bronchi, left and right upper lobe bronchi, left lower lobe bronchus and right truncus intermedius, and so on. We set the speed of the bronchoscope to 10 mm/s, as this is a good approximation of the average speed during bronchoscopy [4]. The sampling rate was set to 40 Hz, according to the frequencies of typical EMT systems such as the NDI Aurora or the Ascension 3D Guidance medSAFE. These data serve as our ground truth bronchoscope paths (marked red in Fig. 1a).

However, as EMT has a significant amount of distortion and jitter, we also added normally distributed noise to the ground truth tracking data to create a more realistic simulation of the bronchoscope paths. According to [14], we set the standard deviation of the noise for a bronchoscope sensor position to 1 and hence for each of its three translational components to $1/\sqrt{3}$. According to the specifications for the orientation accuracy of the NDI tracker and its corresponding orientation distortion for a metal sheet similar to an operating table [15], we set the noise for each of the three Euler angles of a measured camera pose to $\sqrt{(\frac{1}{2} \cdot 0.5)^2 + (\frac{1}{2} \cdot 1.2)^2}/\sqrt{3} = 0.65/\sqrt{3}$.

Once noise is added to the bronchoscope path, assuming normal breathing frequency of 12 breaths per minute, we can transform each sample along the path according to its respiratory information. For this we apply linear interpolation between the 10 respiratory phases to care for inter-phase motion as well as within each deformation field to care for inter-voxel spaces (a resulting bronchoscope path including noise and respiratory motion is marked blue in Fig. 1a). At the same time, to simulate cutaneous sensor measurements for our surrogate data, we select a point at the patient's chest wall and apply the same sampling rate and respiratory motion as for the bronchoscope sensor. However, since the cutaneous sensor only moves within a very limited volume up to about 10 mm and is hence only marginally affected by distortion and systematic error, we only need to take measurement precision into account when simulating error, so according to the specifications of NDI[2] we set its standard deviation for x-, y-, and z-direction (i.e. along the left-right, superior-inferior, and anteroposterior axes) to $(\frac{1}{2} \cdot 0.9)/\sqrt{3} = 0.45/\sqrt{3}$.

Finally, we apply a Euclidean transformation to all simulated sensor measurements, so they lie outside the coordinates of the original CT volume. As now ground truth and simulated data do not overlap any more, we rule out this trivial solution for the optimizer.

For evaluation of the final registration accuracy, we used 37 of the 41 widely distributed landmarks in the POPI model provided by medical experts as ground truth, which are inside the lungs or close to the airways (three of the POPI landmarks are outside these regions, one is an image artifact). These landmarks serve as target points for e.g. biopsies and are not in close proximity to the simulated bronchoscope path used for registration. In our simulation environment, we apply the same noise to these landmarks as to the bronchoscope sensor, deform them according to 10 respiratory phases, and transform them to be outside the original CT volume in the same way as for the bronchoscope path. Figure 1a

[2] http://www.ndigital.com/medical/aurora-techspecs-performance.php

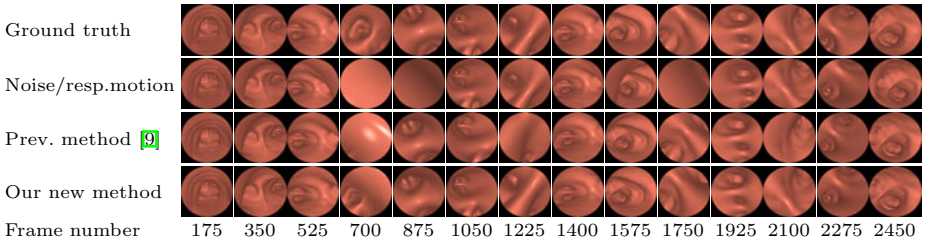


Fig. 2. Exemplary frames of a bronchoscope path, showing ground truth data (first row), after adding noise and respiratory motion (second row), and corresponding renderings after registration with a previous method [9] (third row) and our new method (fourth row)

shows the ground truth landmarks in pink and their corresponding ones in 10 different respiratory phases including noise in black. After obtaining ${}^{\mathrm{CT}}\boldsymbol{T}_{\mathrm{EMT}}$ and $\boldsymbol{t}_{\mathrm{cor}}$ by our two-step registration approach, we can now transform each simulated landmark $\boldsymbol{p}_l$ to $\boldsymbol{p}_{l_{\mathrm{cor}}}$ according to Eq. 2 and compare it to the ground truth data.

Figure 1b shows a bronchoscope path and evaluation landmark positions after applying our registration and motion compensation method. Figure 2 shows exemplary frames of a bronchoscope path (using four branch generations) for ground truth data and after adding noise and respiratory motion as well as the resulting paths after registration with a previously proposed method [9] and our new method. The supplementary video demonstration[3] shows all frames of this path and allows the observer to qualitatively evaluate the performance of our registration method. Table 1 quantitatively compares the registration error of our new method to [9]. We also integrated only the new error term of Eq. 1 into [9] to outline individual accuracy differences.

The runtime of our registration method on a PC with an Intel Xeon X5355 processor was between 1 and 3 seconds and mainly depends on the number of branch generations and hence sample points created along the bronchoscope path.

4 Discussion

As can be seen in Table 1, all methods perform better the more airway generations are covered during registration, so we suggest to perform a registration after exploring as many branches as possible. A one-way analysis of variance (ANOVA) over the 3700 samples (10 paths, 10 respiratory phases, and 37 landmarks) of each case confirms that our new registration method significantly

[3] http://campar.in.tum.de/files/publications/feuerste2010miar.video_path_without_noise.avi and http://campar.in.tum.de/files/publications/feuerste2010miar.video_path_with_noise.avi

Table 1. Registration error in mm over all 10 simulated bronchoscope paths according to absence or presence of noise, number of airway generations and bronchoscope path samples, respiratory phase used to compute the landmark registration error, error term, and utilization of respiratory motion correction

Noise	Gen.	Samples	Phase	Registration error [mm]		
				No respiratory motion correction [9]		New method (Resp. mot. corr. & new error term)
				Old error term [9]	New error term	
w/o	2	1319 ±22	1 2 3 4 5	5.3 4.9 4.6 5.0 5.5	4.6 4.2 3.9 4.5 5.4	4.2 4.3 4.1 4.2 4.5
			10 9 8 7 6	5.2 4.7 4.7 5.2 5.6	4.4 4.1 4.4 5.1 5.7	4.2 4.3 4.5 4.5 4.7
			avg	5.1 ± 2.3	4.6 ± 2.3	4.3 ± 2.4
	3	1815 ±24	1 2 3 4 5	5.0 4.5 4.2 4.6 5.1	3.2 2.6 2.2 3.0 4.1	1.4 1.7 2.4 2.9 3.3
			10 9 8 7 6	4.7 4.2 4.1 4.7 5.2	2.9 2.4 2.7 3.7 4.4	1.8 2.1 2.7 3.1 3.5
			avg	4.6 ± 2.1	3.1 ± 1.7	2.5 ± 1.4
	4	2754 ±25	1 2 3 4 5	3.9 3.3 2.6 3.1 4.0	3.0 2.5 2.0 2.9 4.0	2.0 2.2 2.3 2.5 2.8
			10 9 8 7 6	3.5 2.9 2.8 3.3 4.2	2.5 1.8 2.1 3.3 4.2	2.0 2.2 2.4 2.5 3.0
			avg	3.4 ± 1.8	2.8 ± 1.5	2.4 ± 1.4
w	2	1319 ±22	1 2 3 4 5	5.4 4.8 4.6 4.9 5.6	4.6 4.0 4.0 4.4 5.5	4.6 4.5 4.5 4.5 4.9
			10 9 8 7 6	5.2 4.7 4.6 5.4 5.6	4.4 4.2 4.5 5.4 5.8	4.5 4.7 4.8 5.0 5.1
			avg	5.1 ± 2.4	4.7 ± 2.3	4.7 ± 2.5
	3	1815 ±24	1 2 3 4 5	5.0 4.3 4.2 4.6 5.1	3.4 2.6 2.4 3.0 4.2	3.2 3.2 3.9 4.1 4.6
			10 9 8 7 6	4.8 4.1 3.9 4.9 5.2	3.0 2.6 2.8 3.9 4.5	3.5 3.7 4.0 4.5 4.8
			avg	4.6 ± 2.1	3.2 ± 1.7	4.0 ± 3.7
	4	2754 ±25	1 2 3 4 5	4.1 3.4 2.9 3.0 4.1	3.1 2.5 2.3 2.9 4.1	2.4 2.4 2.8 2.7 3.3
			10 9 8 7 6	3.7 3.1 3.0 3.5 4.2	2.7 2.1 2.4 3.5 4.4	2.5 2.6 2.8 3.0 3.4
			avg	3.5 ± 1.8	3.0 ± 1.5	2.8 ± 1.6

outperforms the previous method [9] with a p-value of 0.000 in all cases. This is because it successfully utilizes the actual curved shape of the airways in contrast to [9] only using straight line segments. Furthermore, our new method first tries to find the bronchoscope sensor phase best matching the CT data and based on that linearly adjusts the respiratory motion offset. The positive effect of this adjustment can be seen e.g. in frames 700, 1225, and 2100 of Fig. 2.

In the presence of noise, our method showed some outliers when registration was performed using 3 airway generations, leading to an error of 4.0 ± 3.7 mm. As we attribute this behavior to too noisy cutaneous sensor data, we repeated our experiments only adding noise to the bronchoscope sensor, but not to the cutaneous sensor, simulating a perfect respiratory signal. Since this reduced the registration error for 3 airway generations from 4.0 ± 3.7 mm to 2.8 ± 1.5 mm and for 4 generations from 2.8 ± 1.6 mm to 2.6 ± 1.3 mm, we are now considering using a more reliable technique for measuring respiratory motion such as an elastic belt placed around the abdomen, which is a well established means for gated 4D CT and MR, SPECT, and PET-CT imaging as well as radiation therapy, or, if a belt is not available, a method to denoise the surrogate data and match it to a realistic respiratory curve.

While for the previous method error is distributed over all 10 breathing phases, for our method it is significantly lower, but at the same time increases between end-inhalation (phase 1) and end-exhalation (phase 6). This confirms that our registration method approximates patient breathing well, but still not to its full extent. One may hence think of matching a (readily available) dense deformation field of another patient to the current patient to obtain a more accurate estimation of dense respiratory displacements instead of using a linearly scaled

translational offset for breathing compensation. However, our current results using a simple and fast approach for automated registration are very promising. As soon as permission is granted by our institution, we will also evaluate our method *in vivo*.

In our experiments we also tested other methods to compensate for respiratory motion such as scaling the tracking data (as lung motion is bigger at the diaphragm than towards the superior) and rotating all branches around their branching points (as their motion is not completely linear). However, a simple translation gave the best results. When optimizing translation and scaling at the same time, the registration error for the case of four airway generations and noisy data increased to 2.9 to 3.9 mm (compared to 2.8 mm), depending on the choice of which axes to scale and the position of the scaling center along the craniocaudal axis. We attribute this to an overfitting of the tracking data to the bronchial tree due to additional degrees of freedom. When optimizing scaling only, the registration error was 4.0 to 4.3 mm.

The automatic approach for phase determination using only rigid registrations at first glance may sound paradoxical under the assumption of a linear respiratory motion model. However, it works well, because the translation assumed for respiratory motion compensation is only an approximation of the real deformation, which is more complex. Under this complex deformation it should be clear that a CT rigidly registered to tracking data from the same phase should fit better than to tracking data from a different deformed phase. Our initial rigid registration only considers two subsets of the tracking data, and only yields a single transformation. The second registration then includes all tracking data and yields an additional time-dependent translation which approximates the respiratory motion.

Finally, our method currently only addresses respiratory motion. We are planning to also add methods for the handling of other kinds of patient motion, for instance one that is able to exclude outliers from surrogate data caused by e.g. coughing.

5 Conclusion

This paper successfully reduces registration error caused by respiratory motion and insufficient utilization of the correct airway shape during optimization. Using the POPI model, we are able to provide a very realistic respiratory motion simulation with real patient data, making a laborious *in vivo* experiment unnecessary for a first thorough evaluation. Our evaluation shows that our new automated marker-free registration method is significantly more accurate than the current state of the art [9]. Even for noisy data, as it is the case for EMT data, our method performs robust and reduces the registration error from 3.5 mm to 2.8 mm.

References

1. Eberhardt, R., Anantham, D., Herth, F., Feller-Kopman, D., Ernst, A.: Electromagnetic navigation diagnostic bronchoscopy in peripheral lung lesions. Chest 131(6), 1800–1805 (2007)
2. Bertoletti, L., Robert, A., Cottier, M., Chambonniere, M.L., Vergnon, J.M.: Accuracy and feasibility of electromagnetic navigated bronchoscopy under nitrous oxide sedation for pulmonary peripheral opacities: An outpatient study. Respiration 78(3), 293–300 (2009)
3. Bricault, I., Ferretti, G., Cinquin, P.: Registration of real and CT-derived virtual bronchoscopic images to assist transbronchial biopsy. IEEE Trans. Med. Imag. 17(5), 703–714 (1998)
4. Rai, L., Helferty, J.P., Higgins, W.E.: Combined video tracking and image-video registration for continuous bronchoscopic guidance. Int. J. CARS 3, 315–329 (2008)
5. Gergel, I., dos Santos, T.R., Tetzlaff, R., Maier-Hein, L., Meinzer, H.P., Wegner, I.: Particle filtering for respiratory motion compensation during navigated bronchoscopy. In: SPIE Medical Imaging (2010)
6. Soper, T.D., Haynor, D.R., Glenny, R.W., Seibel, E.J.: In vivo validation of a hybrid tracking system for navigation of an ultrathin bronchoscope within peripheral airways. IEEE Trans. Biomed. Eng. 57(3), 736–745 (2010)
7. Deguchi, D., Ishitani, K., Kitasaka, T., Mori, K., Suenaga, Y., Takabatake, H., Mori, M., Natori, H.: A method for bronchoscope tracking using position sensor without fiducial markers. In: SPIE Medical Imaging (2007)
8. Klein, T., Traub, J., Hautmann, H., Ahmadian, A., Navab, N.: Fiducial-free registration procedure for navigated bronchoscopy. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 475–482. Springer, Heidelberg (2007)
9. Mori, K., Deguchi, D., Kitasaka, T., Suenaga, Y., Hasegawa, Y., Imaizumi, K., Takabatake, H.: Improvement of accuracy of marker-free bronchoscope tracking using electromagnetic tracker based on bronchial branch information. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 535–542. Springer, Heidelberg (2008)
10. Borgert, J., Krüger, S., Timinger, H., Krücker, J., Glossop, N., Durrani, A., Viswanathan, A., Wood, B.J.: Respiratory motion compensation with tracked internal and external sensors during CT-guided procedures. Comput. Aided Surg. 11(3), 119–125 (2006)
11. Vanden Berghen, F., Bersini, H.: CONDOR, a new parallel, constrained extension of powell's UOBYQA algorithm: Experimental results and comparison with the DFO algorithm. J. Comput. Appl. Math. 181, 157–175 (2005)
12. Vandemeulebroucke, J., Sarrut, D., Clarysse, P.: The POPI-model, a point-validated pixel-based breathing thorax model. In: ICCR (2007)
13. Feuerstein, M., Kitasaka, T., Mori, K.: Automated anatomical likelihood driven extraction and branching detection of aortic arch in 3-D chest CT. In: Second International Workshop on Pulmonary Image Analysis, pp. 49–60 (2009)
14. Wegner, I., Tetzlaff, R., Biederer, J., Wolf, I., Meinzer, H.P.: An evaluation environment for respiratory motion compensation in navigated bronchoscopy. In: SPIE Medical Imaging (2008)
15. Kirsch, S.R., Schilling, C., Brunner, G.: Assessment of metallic distortions of an electromagnetic tracking system. In: SPIE Medical Imaging (2006)

Computational Decision Support for Percutaneous Aortic Valve Implantation

Ingmar Voigt[1,3,⋆], Razvan Ioan Ionasec[2,4], Bogdan Georgescu[2], Jan Boese[5], Gernot Brockmann[6], Joachim Hornegger[5], and Dorin Comaniciu[2]

[1] Software and Engineering, Siemens Corporate Technology, Erlangen, Germany
ingmar.voigt.ext@siemens.com
[2] Integrated Data Systems, Siemens Corporate Research, Princeton, USA
[3] Pattern Recognition Lab, Friedrich-Alexander-University, Erlangen, Germany
[4] Computer Aided Medical Procedures, Technical University Munich, Germany
[5] Siemens Healthcare, Angiography & X-Ray-Systems, Forcheim, Germany
[6] German Heart Center Munich

Abstract. Valve replacement is the most common therapy for diseased aortic valves. Percutaneous approaches are becoming increasingly popular, due to reduced procedural complications and lower follow-up rates. Still there is a lack of efficient tools for valve quantification and preoperative simulation of replacement and repair procedures. Thus the success of the intervention relies to a large portion on experience and skills of the operator. In this paper we propose a novel framework for preoperative planning, intraoperative guidance and post-operative assessment of percutaneous aortic valve replacement procedures with stent mounted devices. A comprehensive model of the aortic valvular complex including aortic valve and aorta ascendens is estimated with fast and robust learning-based techniques from cardiac CT images. Consequently our model is used to perform a in-silico delivery of the valve implant based on deformable simplex meshes and geometrical constraints. The predictive power of the model-based in-silico valve replacement was validated on 3D cardiac CT data from 20 patients through comparison of preoperative prediction against postoperatively imaged real device. In our experiments the method performed with an average accuracy of 2.18 mm and a speed of 55 seconds. To the best of our knowledge, this is the first time a computational framework is validated using real pre- and postoperative patient data.

1 Introduction

Percutaneous aortic valve implantation (PAVI) has the potential to revolutionize the treatment of aortic valve disease, offering a less invasive alternative to open heart surgery. PAVI is already emerging as a feasible treatment for patients with high-surgical risk [1], over 30% of the symptomatic cases, and will account for 41.1% of the procedures by 2012 (Millennium Research Group 2008)

⋆ Correspondence author.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 247–256, 2010.

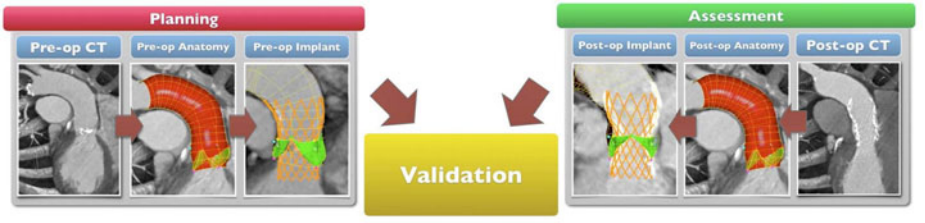


Fig. 1. Schematic description of the proposed PAVI computational decision support workflow

[2]. The prosthetic implants are delivered through catheters using transvenous, transarterial or transapical techniques, while clinicians do not have direct view and access to the affected valve and surrounding anatomies.

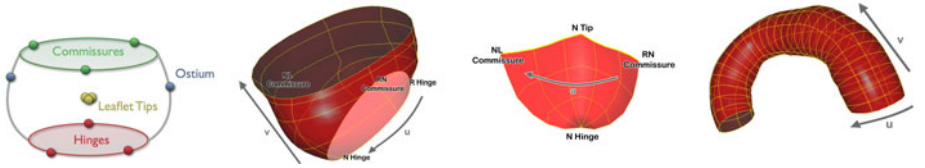
Hence, critical decisions such as, type of procedure, implant type and sizing, deployment location and timing, and treatment assessment, are exclusively based on imaging techniques [3]. A misplaced implant can block the coronary ostia inducing a life threatening ischemic condition. Suboptimal deployment location can result in poor hemodynamic performance with severe paravalvular leakages and/or high gradients and suboptimal effective orifice. Wrong implant sizing may require re-operation or can damage the vessel tissue and cause catastrophic events as arterial dissection or rupture. Therefore advanced image analysis and computational models for precise planning, procedure guidance, and outcome assessment, may significantly improve percutaneous valve implantation techniques.

In this paper, we propose a computational framework for percutaneous aortic valve implantation, which supports decisions throughout the clinical workflow and is summarized in Sec. 2. Modeling of the aortic valve and ascending aorta and patient-specific estimation from pre- and post- operative cardiac CT images is described in Sec. 3. Sec. 4 presents the computational environment, which allows for in-silico valve implantation for evaluation and prediction of procedure success under various treatment scenarios. Comprehensive validation and performance evaluation is given in Sec. 5 by comparing the simulation results from preoperative data with the real device imaged in the postoperative data.

2 Computational Decision Support for PAVI

The proposed PAVI computational decision support workflow is illustrated in Fig. 1:

Pre-operative workflow: 1) Pre-operative cardiac CT volume acquisition for procedure planning purposes 2) Patient-Specific anatomical model estimation and automatic quantification for valve assessment and patient selection 3) In-silico valve implantation under various interventional procedure conditions for identification of optimal device type, size and deployment location as well as treatment outcome prediction until optimal predicted performance is observed.



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

Fig. 2. Aortic valve and ascending aortic root model. (a) shows a generic model of the aortic valve including nine anatomical landmarks. (b) shows our point distribution model of the aortic root. (c) presents the aorta leaflets model - the N leaflet is depicted. (d) demonstrates the ascending aortic root model. (e) represents the full model with the corresponding anatomical parameterization.

Post-operative workflow: 4) Post-operative cardiac CT volume acquisition for treatment evaluation 5) Patient-Specific anatomical model estimation for quantitative anatomical assessment 6) Patient-Specific deployed device estimation for quantitative implant assessment.

3 Patient-Specific Anatomical Modeling and Estimation

This section summarizes the anatomical model of the aortic valve and ascending aorta as well as the patient-specific estimation of its parameters from imaging data as in [4].

3.1 Aortic Valve and Ascending Aortic Root Modeling

The aortic root provides the supporting structures for the leaflets of the aortic valve and forms the bridge between the left ventricle and the ascending aorta. The root extends from the basal attachments of the leaflets, defined by the L (left) / R (right) / N (none) Hinges, to the sinutubular junction. The L / R / N aortic leaflets, are attached to the root on semilunar structures. Valve leaflets can be thought of as shirt pockets, with one edge stitched to the shirt and one free of attachment with is center marked by the L / R / N leaflet tips. These attachment structures interlink at the level of the sinutubular junction forming the LR / RN / NL commissures. The employed model represents the complete anatomy of the aortic valve and ascending aorta, which includes the aortic root, left / right / none aortic leaflets, ascending aorta and 11 anatomical landmarks.

Anatomical Landmarks: Represented by three-dimensional points in the Euclidean space, the considered anatomical landmarks are: L / R / N Hinges, LR / RN / NL commissures, L / R / N leaflet tips, and L / R coronary ostia.

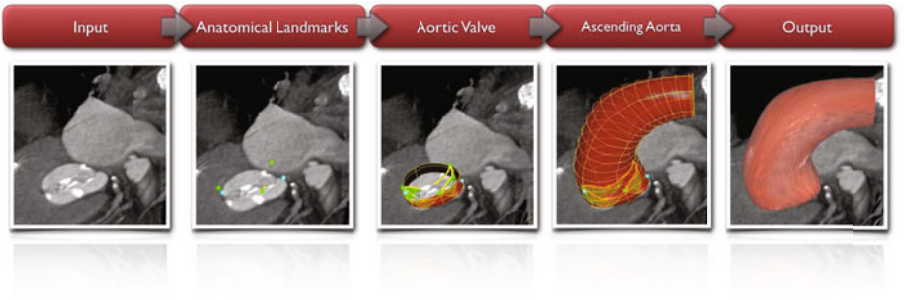


Fig. 3. A survey of our hierarchical model estimation schema

Aortic valve root and leaflets: The aortic valve root is constrained by the commissures, hinges and ostia and represented as a tubular surface mesh. The mesh is aligned with the aortic circumferential u and ascending directions v and includes 36×10 vertices. The left / right / none aortic leaflets, are modeled as hyperbolic paraboloids on a grid of 11×7 vertices. Each leaflet is defined by one hinge, two commissures and one leaflet tip.

Ascending aortic root: The ascending aorta emerges from the aortic root and incorporates a variable length. The anatomy is composed of a fixed number of circumferential coordinates $u = 36$ and a variable number of coordinates along the ascending direction v. The first ring starts at from the commissures.

3.2 Patient-Specific Model Estimation

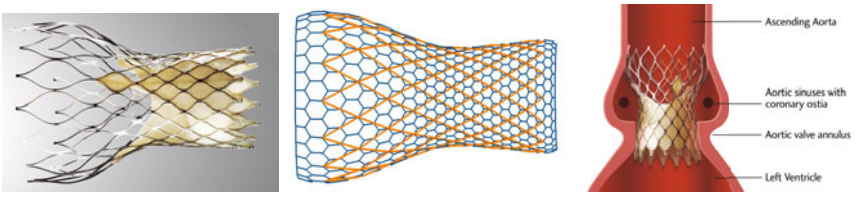
The patient-specific parameters of the aortic valve and ascending aorta model described in Sec. 3.1 are estimated from volumetric images using a robust learning-based algorithm as in [5]. The a posteriori probability $p(M|I)$ of the model M given the image data I, is hierarchically estimated within the Marginal Space Learning (MSL) [6] framework. Detectors are successively trained using the Probabilistic Boosting Tree (PBT) [7] with Haar and Steerable features, and consequently applied to estimate the anatomical landmarks and structures from cardiac CT volumes as illustrated in Fig. 3. For further details the reader is referred to [4].

4 Device Modeling and In-Silico Deployment

4.1 Stent Model

A library of virtual devices/implants was created based on manufacturers' description to incorporate realistic geometrical properties. In this work two models of the CoreValve Revalving System by Medtronic (Minneapolis, MN, USA) are treated, namely the models CRS-P3-640 and CRS-P3-943 (Fig. 4(a)). The implant consists of 165 cells formed by the struts. The two models have length of 53

and 55 mm and diameters at the inflow, middle and outflow levels of 26, 22, 40 and 29, 24, 43 mm respectively. The Xenograft artificial valve consist of porcine pericardial tissue, out of which the leaflets are manufactured and mounted to the implant's stent. The library can be easily extended with future devices using the methods described in the following. The device is modeled out of two parts: a geometric representation, which precisely mimics the exact geometry of the device, the so-called *stent mesh*, and a second superimposed 2-simplex mesh, named in the following *computational mesh*, which is used for computation and to guide the expanding deformation [8,9]. Fig. 4(b) depicts the topological relationship between the computational mesh and the stent mesh, which is composed of struts connecting a subset of points of the computational mesh. In order to infer the geometrical properties of the stent model various dimension were measured from stereolithographic scans of the modeled implants. These are the strut lengths, the characteristic angles in each cell and the device's circumferences at each level, where each level is defined by the strut joints.



(a) (b) (c)

Fig. 4. (a) CoreValve implant, (b) long axis cross section of stent mesh (orange) with superimposed computational mesh (blue) and (c) CoreValve implant with sketch of target anatomy. (Sources a & c: http://www.medtronic.com)

4.2 Virtual Stent Deployment

To simulate valve replacement under various conditions, different devices are chosen from the library and virtually deployed under different parameters, into the previously extracted patient-specific model of the affected valve. The expansion of the device is modeled by balancing external and internal forces as encountered in the actual procedure, using iterative optimization methods. Following the works of Larrabide et. al. and Montagnat et. al. [8,9], the expansion is described by a finite difference discretization of a second order differential equation:

$$\boldsymbol{p_i^{n+1}} = \boldsymbol{p_i^n} + (1-\gamma)(\boldsymbol{p_i^n} + \boldsymbol{p_i^{n-1}}) + f_{int}(\boldsymbol{p_i^n}) + f_{ext}(\boldsymbol{p_i^n}) + f_{reg}(\boldsymbol{p_i^n}) \quad (1)$$

where $\boldsymbol{p_i}$ is a point on the computational mesh, n is the iteration number, f_{ext}, f_{int} and f_{reg} external, internal and regularizing forces and the weighting parameter γ. Fig. 5 shows a visual description of each of the forces. An outline of the algorithm is given in Fig. 6. The internal forces $f_{int}(\boldsymbol{p_i^n}) =$

$f_{length}(\boldsymbol{p_i^n}) + f_{angle}(\boldsymbol{p_i^n}) + f_{circ}(\boldsymbol{p_i^n})$ model the intrinsic properties of the stent and enforce deformation along it's surface normals and long axis as the device is self-expandable. Hence they are parameterized by strut lengths, characteristic angles and device circumferences, which were measured from the expanded template. Accordingly, these forces are adapted, such that the implant attempts to achieve the targeted dimensions, and they induce different expanding pressures at different levels. Particularly $f_{circ}(\boldsymbol{p_i^n}) = \boldsymbol{n}_i(c_k - \sum_{\forall j \in \mathcal{N}_k} ||\boldsymbol{p_j^n} - \boldsymbol{p_{j+1}^n}||)/2\pi$ pushes the points $\boldsymbol{p_i^n} \in \mathcal{N}_k$ along the surface normal $\boldsymbol{n}_i$ to satisfy the reference circumference c_k of the stent shape, where $\mathcal{N}_k$ is the set of strut joints at a level k. It is important to note, that f_{circ} does not enforce the stent diameter directly but the stent circumference instead to account for expansion into arbitrary shaped vessel geometries, which have typically non-circular cross sections. f_{length} and f_{angle} enforce the strut lengths and characteristic angles observed in the expanded shape [8]. The external forces $f_{ext}(\boldsymbol{p_i})$ model the interaction of stent and aortic valve and aorta tissue, and guide the implant deformation by balancing the internal device forces: $f_{ext}(\boldsymbol{p_i}) = -\boldsymbol{n}_i(\boldsymbol{n}_i \cdot f_{int}(\boldsymbol{p_i}))(||\boldsymbol{p_i^n} - \boldsymbol{c}_k||/||\boldsymbol{v} - \boldsymbol{c}_k||)$ with stent centroid $\boldsymbol{c}_k$ at level k and the intersection point $\boldsymbol{v}$ of normal and vessel surface. The regularizing forces f_{reg} are solely defined on the computational mesh to provide smoothness as described in [9]. As mentioned above the method focusses on self-expanding implants, which inherently exercise forces of minor amplitudes onto the surrounding vessel tissue. Therefore we argue, that the resulting minor deformations can be neglected.

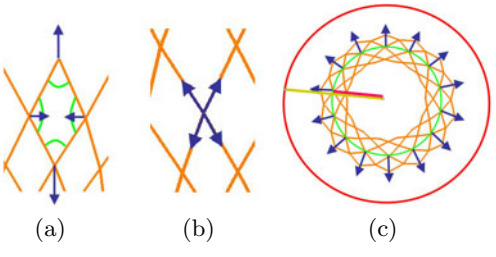


Fig. 5. Forces acting on the model on deployment to converge to the observed geometric properties: (a) f_{angle} enforces the charateristic angles at the strut joints (green), (b) f_{length} maintains the strut lengths. (c) f_{circ} enforces the circumference (green), while f_{ext} dampens and eliminates the all forces acting along the stent mesh normal wheighted by the fraction of distances of strut joint and vessel wall (red) to the stent centroid (magenta/yellow). Please note that (c) shows a short axis cross section of the stent mesh.

5 Experimental Results

The validation of the proposed framework is divided in two experiments. First we present results on the performance of the automatic patient-specific model estimation from pre- and post- cardiac CT data, as well as the quantitative variation

Input:

- Patient-specific model of aortic valve and aorta ascendens
- implant placement position and orientation

Output: Deployed Implant
Execute:

- create computational mesh and stent mesh with constant radius of 1 mm at manually selected placement position, oriented along the aortic root centerline
- **repeat:**
  - for each point $\boldsymbol{p_i^n}$ on the *computational mesh*, calculate $f_{reg}(\boldsymbol{p_i^n})$, $f_{angle}(\boldsymbol{p_i^n})$, $f_{length}(\boldsymbol{p_i^n})$, $f_{circ}(\boldsymbol{p_i^n})$ and $f_{ext}(\boldsymbol{p_i^n})$
  - for each $\boldsymbol{p_i^n}$, compute $\boldsymbol{p_i^{n+1}}$ according to Eq. 1
  - if mean point displacement on the *stent mesh* $< \epsilon$, convergence achieved; stop execution

Fig. 6. The outline of our virtual stent deployment algorithm

between pre and postop ground truth anatomies, which is relevant for the subsequent virtual imlant deployment. Second we validate the proposed in-silico implantation, by comparing predicted valve deployment, using pre-operative data, with real deployment from post-operative data.

5.1 Validation of Patient-Specific Anatomical Modeling and Parameter Estimation

The data set used for patient-specific model estimation consists of 63 multi-phase (10 frames per cycle) cardiac CT and 21 single-phase cardiac CT acquisitions, which sums up to 651 CT volumes. Scans are acquired from different patients with various cardiovascular diseases (including ascending aortic root aneurysm, regugitation, calcific stenosis and bicuspid aortic valves), using different protocols, resulting in volumes with 80 to 350 slices and 153x153 up to 512x512 voxel grid resolution and 0.28mm to 2.0mm spatial resolution. Each data set is associated with an expert annotation used as ground-truth.

For the automatic patient-specific anatomical model estimation a combined accuracy of 1.45mm is obtained in 30sec on a standard desktop machine (Intel Xeon 2.66Ghz, 2GB RAM) for both pre- and post-operative volumes. Performance is reported on test data, which represents randomized 20% of the complete dataset, while the remaining 80% were used for training.

Due to different factors, a bias between the pre- and post-operative anatomical ground truth models can be expected. These are cardiac phase shifts and image noise but also deformation of the aortic vessel wall due to stent deployment, where the latter was assumed to be sufficiently small to be neglected in the deployment algorithm (Sec. 4.2). Therefore we quantified the differences for each pair of corresponding anatomical models obtained from a subset of 20 patients with pre- and postoperative image data. The quantitative results in Table 1

Table 1. Deviation of pre- and postoperative ground truth anatomical models: Differences in diameter at sinutublar junction, valsava sinuses and aortic annulus are given in absolute values as well as relative to the postoperative measurement. Values of Mean and standard deviation are provided as well as 80-percentile and maximum.

	absolute (mm)			relative (%)		
measurement	mean (std)	80%	max	mean (std)	80%	max
sinutublar junction	2.3 (1.7)	3.7	5.7	6.46 (4.6)	10.5	14.9
valsava sinuses	1.1 (0.9)	1.7	4.1	3.49 (2.6)	5.2	9.98
annulus	1.5 (1.2)	2.5	5.2	5.06 (3.2)	7.7	14.3
point-to-mesh distance	1.6 (0.98)	2.4	2.8	-	-	-

support the validity of our assumption, showing a mean relative deviation of up to 6.46% between pre- and post-operative anatomies.

5.2 Validation of In-Silico Implant Deployment

The validation of the in-silico implant deployment is performed on 20 patients with pre- and post-operative cardiac CT images, affected by various diseases such as calcific stenosis as mentioned in the previous section. It is important to note, that for this purpose the preoperative prediction result is compared with the real device imaged in the postoperative data, where the latter serves as a ground truth for this experiment.

The implant is virtually deployed into the associated anatomical model of the preoperative volume using the algorithm described in Sec 4.2. In the postoperative volume the ground truth implant is manually placed and fit to the imaged stent, which is well visible in image data, using a semi-automatic method based on the thin-plate-spline transformation. In the envisioned target application the optimal deployment location and orientation is manually selected by the clinician. For validation purposes this is indirectly available and has to be inferred by registering pre- and post-operative anatomical models. A selection of virtually deployed vs. their corresponding ground truth stents is depicted in Fig. 7. The performance is reported in Table 2 in terms of internal precision, by comparing only the virtual and real implants shape in isolation via symmetric point-to-point distance, and external precision. The latter means to compare the virtual and real implants position relative to clinically relevant locations, in order to account for the potentially critical conditions due to wrong implant sizing and placement such as blockage of coronary ostia and more importantly paravalvular leakages at the annular level as mentioned in Sec 1. This is done by computing the differences of the pre- and postoperatively measured distances from annulus ring and coronary ostia to the closest stent point respectively.

In the clinical context, the required accuracy is proportional to the tolerance between therapeutical alternatives. Considering the diameter differences of 3mm (at the annular level) of the Medtronic CoreValve implants (Sec 4.1), the system provides a sufficient approximation in at least 80% of the cases for prevention of

Table 2. Accuracy of in-silico valve deployment quantified from preop deployment prediction vs. postop ground truth stent and measured in mm: besides point-to-point distance, accuracy relative to the anatomies was estimated from the differences in distances between aortic valve annulus and coronary ostia and implant. Values of mean and standard deviation are provided as well as 80-percentile and maximum.

	mean (std)	80%	max
stent point-to-point	2.18 (1.77)	2.4	8.45
annulus	0.7 (0.73)	1.4	2.14
L coronary ostium	1.42 (1.51)	2.16	4.75
R coronary ostium	1.55 (1.24)	2.02	4.27

paravalvular leakages, with an external accuracy of up to 1.4mm at the annular level. The algorithm performed at an average speed of 55sec on a standard desktop machine (Intel Xeon 2.66Ghz, 2GB RAM). Thus our framework enables for fast and efficient preoperative planning and risk minimization by finding the best implant type, size and deployment location and orientation via varying these parameters until optimal predicted performance is observed.

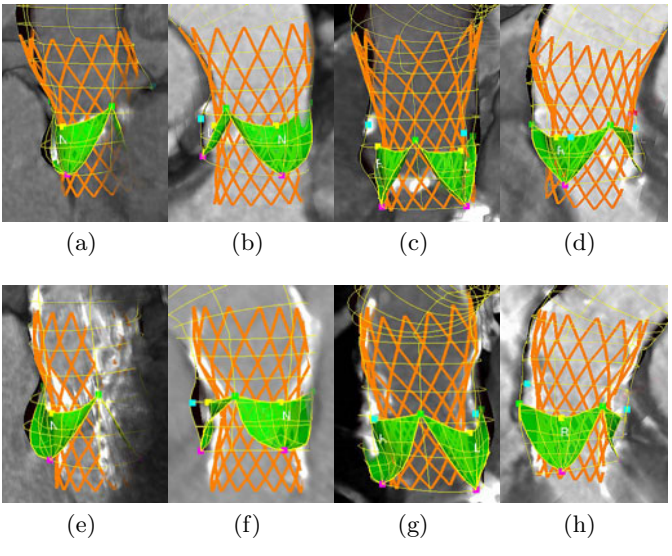


Fig. 7. Example results of preoperative virtual stent deployment (a-d) vs. postoperative ground truth stents (e-h) overlayed with with the anatomical models. Note the deviation of the virtually deployed stent around the sinutubular junction (upper end) in contrast to the close approximation at sinus and annular level, which is to due to the fact, that internal stiffness of the stent configuration is not modeled yet.

6 Discussion

In this paper a framework for computational decision support for percutaneous aortic valve implantation was presented. A fast and robust estimation of an anatomical model enables for precise modeling of the patient-specific morphology and is consequently used for in-silico implant deployment. The approach was validated with pre- and post-operative data sets from 20 patients and shows reasonable accuracy within the variation in appearance given by image and motion artifacts. To the best of our knowledge, this is the first time a computational framework is validated using real pre- and postoperative patient data. The framework is targeted for fast and efficient preoperative planning with a library of different implants, intraoperative guidance and postoperative assessment of interventional outcome. It may have impact on the cardiology of the future and improve the OR towards increased transparency.

References

1. Grube, E., Laborde, J., Gerckens, U., Felderhoff, T., Sauren, B., Buellesfeld, L., Mueller, R., Menichelli, M., Schmidt, T., Zickmann, B., Iversen, S., Stone, G.W.: Percutaneous implantation of the corevalve self-expanding valve prosthesis in high-risk patients with aortic valve disease. Circulation (114), 1616–1624 (2006)
2. Lloyd-Jones, D., et al.: Heart disease and stroke statistics - 2009 update. Circulation (2009)
3. Otto, C., Bonow, R.: Valvular Heart Disease: A Companion to Braunwald's Heart Disease. Saunders (2009)
4. Grbic, S., Ionasec, R.I., Zheng, Y., Zaeuner, D., Georgescu, B., Comaniciu, D.: Aortic valve and ascending aortic root modeling from 3d and 3d+t ct. In: SPIE Medical Imaging, San Diego, USA (February 2010)
5. Ionasec, R.I., Voigt, I., Georgescu, B., Houle, H., Hornegger, J., Navab, N., Comaniciu, D.: Modeling and assessment of the aortic-mitral valve coupling from 4d tee and ct. In: Proc. MICCAI, London, USA (September 2009)
6. Zheng, Y., Georgescu, B., Ling, H., Zhou, S.K., Scheuering, M., Comaniciu, D.: Constrained marginal space learning for efficient 3d anatomical structure detection in medical images. In: CVPR, pp. 194–201 (2009)
7. Tu, Z.: Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In: Proc. ICCV, Washington, DC, USA, pp. 1589–1596 (2005)
8. Larrabide, I., Radaelli, A., Frangi, A.F.: Fast virtual stenting with deformable meshes: Application to intracranial aneurysms. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 790–797. Springer, Heidelberg (2008)
9. Montagnat, J., Delingette, H.: 4d deformable models with temporal constraints: application to 4d cardiac image segmentation. Medical Image Analysis 9(1), 87–100 (2005)

Relative Error: An Approach for *in vivo* Characterization of Electromagnetic Tracking Errors and Confidence Intervals

Mohammad Matinfar, Vijay Parthasarathy, Raymond Chan, and Ameet Jain

Department of Interventional Guidance Technology
Philips Research North America, Briarcliff, NY
ameet.jain@philips.com

Abstract. The sensitivity of electromagnetic (EM) tracking to electromagnetic field distorters in a clinical environment is well known. Correction of the EM field distortion is challenging given the ill-posed nature of the problem and the typically sparse set of observations available in computing an inverse solution. Furthermore, interventional environments dynamically change during a procedure making pre-procedural calibration measurements difficult to apply. In this paper, we present the use of relative measurement errors derived from known geometric features of interventional instruments to detect reliable regions of operation. Relative error can be estimated from consecutive measurements of EM sensor locations without the need for additional information beyond the geometry of an interventional tool or *in vivo* phantom. We demonstrate the use of a calibration phantom that allows us to determine the statistical relationship between "relative" errors and "absolute" errors. Using this statistical relationship, we develop a method to compute an estimate of the absolute error, which provides real time information about measurement confidence.

1 Introduction

Many clinical procedures require the tracking of medical instrumentation within the patient's body. In these applications, optical tracking systems cannot be used because of the line of sight (LOS) constraint. EM tracking systems allows for tracking without the LOS constraint. EM tracking systems perform relatively well in clean, non-ferromagnetic environments, but significant performance degradation has been reported when there are source of distortion near the field generator or sensor coils [1], [2]. The measurement errors have been reported to be several millimeters in position and several degrees in orientation [3]. The magnitude of errors observed in an interventional setting imposes limitations on clinical workflow. For example, in cardiac electrophysiology procedures for atrial fibrillation ablation, electroanatomical maps derived from EM-tracked catheters are known to be relevant only when the maps are acquired with the X-ray detector placed as far away as possible from the patient thorax. However, this limits the use of X-ray imaging which is the workhorse for visualization and

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 257–266, 2010.

guidance in the interventional suites. Whenever the X-ray C-arm is moved to acquire new imaging data, the EM environment changes and the electrophysiologist has to repeat electroanatomical maps in order to ensure that ablations are not being performed based on distorted measurements.

Different protocols have been proposed to evaluate the accuracy of these systems in different clinical environments [4], [5] and [6]. They are mainly referred in literature as methods for EM tracking calibration since they tend to perform a onetime calibration and then by using an interpolation algorithm to correct the measurements. In practice, medical intervention suites are dynamically changing and operating instruments that produce EM distortions actively move within the workspace of interest. These instruments have different sizes and different material composition and some instruments have significant distorting effects such as X-ray gantries, CT scanners, etc. These distortions vary with instrument positioning and create a dynamic environment where makes it almost impossible to predict the distortion error in great detail. The main issue with calibration methods is the high dimensionality of the error space and sparsity of calibration measurement points, making this a classic ill-posed inverse problem. We believe that no matter how accurately error is modelled, it is not possible to use a single unwarping function to correct for errors in all environments. These methods are time-consuming and can only compensate for static errors in a single environment when the number of calibration sample points is sufficiently large [7].

Other researchers have proposed hybrid methods for error compensation [8] and [9]. By fusing optical and EM tracking systems, the EM measurement errors are corrected. Specifically in [8], authors used the EM relative error measurements along with a look-up-table resulted from optical and EM systems registration to improve the tracking accuracy. In our approach, we show that for a variety of environments, a correlation exists between absolute error (AE) and relative error (RE). This statistical model is established by collecting a large number of data and finding the probability mapping between AE and RE. The model provides an approach for estimation of "absolute error" without requiring any reference measurements from another tracking system such as optical or a robot. No such statistical histogram has been addressed in literature and is the key difference from prior art. This model is then being validated and assessed and has been shown that in more than 75% of experiments, we are able to classify the AE with an error of 0.46mm in a 2mm error predication.

2 Relative Error vs. Absolute Error

Conventionally, electromagnetic tracking error is defined as the registration error between the EM tracker measurement and a reference navigation system, such as a robot or an optical tracking system.This error is used to draw an error mapping and analyze the workspace of a tool to examine the behavior of the error. It provides the evaluation of the workspace necessary to guide the tool in the zones that the error is minimal. The mapping of absolute errors in dynamic environments is the best way of analyzing the workspace, however, it requires

a reference tracking system which is not practical in clinical application. In this paper, we introduce the concept of relative error to estimate the fidelity of the working EM space. We define relative error as the difference between the absolute errors in two sensors whose absolute errors are highly correlated with each other. In this approach, error is computed for a reference point and consecutive measurement errors are displayed relative to the reference point in order to characterize the accuracy in near-real-time. The concept of relative error is shown intuitively in Fig. 1. The real (known) position of a sensors differs from its tracked position by absolute error. While the absolute errors might be high, the relative error between the two sensor positions is lesser because the error fields are correlated and fairly homogenous.

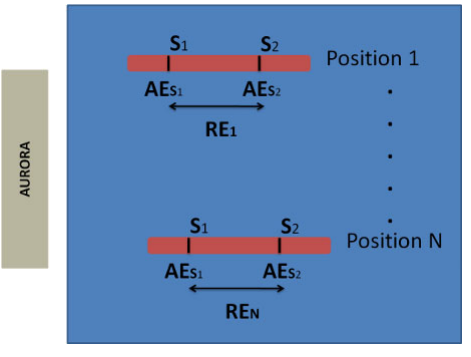


Fig. 1. Relation between relative and absolute error; two EM sensors are placed in a tool where moves inside the workspace of the EM tracker

In order to establish a statistical relation between the relative error and absolute error, we consider the measurements from sensor 1 and sensor 2 as two random variables. Relative error (RE) is the difference between the absolute errors (AE) and defined as:

$$RE = ||EM_{s1} - EM_{s2}| - |Ref_{s1} - Ref_{s2}|| \tag{1}$$

where Ref_{si} and EM_{si} are the measurements of sensor i position in the reference and EM coordinate systems. The term $|Ref_{s1} - Ref_{s2}|$ is the distance between two sensors and is given by the design. In other words, relative error is the difference between the measured distance of sensors s_1 and s_2 and the ground truth distance, either given by design or computed during calibration (as described in Section 3). The absolute error AE is defined as:

$$AE_i = |EM_i - Ref_i| \tag{2}$$

In this work, we compute absolute error from measurements of sensor one (AE_{s1}). It is possible to take the average of both sensor measurement, $(AE_{s1} + AE_{s2})/2$, and assign it to AE, but we found out that the statistics does not differ. It is

important to note that because of high-dimensionality of absolute error, RE is a well defined measure to recover all the variations in AE. Even a uniform shift (bias) in a localized area results in a non-uniform variation of AE rather than a constant shift and therefore it is detected by RE.

The observed relative error can be mapped to the corresponding absolute error to estimate the statistics of absolute error in any location. Given the value of RE at each location the minimum-mean square error estimator of absolute value of error given the relative error measurement is the expected value of the conditional probability. Therefore,

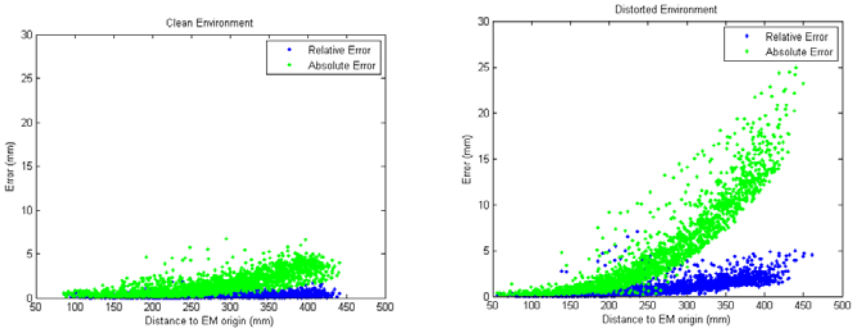
$$\widehat{AE} = E[AE|RE] = \int (ae) f(AE|RE)\, d(ae) \tag{3}$$

where f is the conditional probability density function of absolute error for a known relative error. Equation 3 suggests that if the 2-D joint probability distribution is empirically measured, then statistics of the absolute errors can be estimated using the relative errors. This is done statistically by collecting large samples of data and observing the relationship between AE and RE.

3 Experimental Setup and Error Measurement

In order to validate the concept of relative error, we have designed a cubical phantom that accommodates 6 sensors, with 2 sensors in X, Y and Z axes. Fig. 2 depicts one embodiment of the phantom. The sensors are located 10mm from each other in each axis to satisfy the distance requirement from correlation factor and at the same time avoid interfering with each other. In an initial configuration phase, this phantom is calibrated. By moving the phantom in a clean, distortion-free environment, data is collected for the 2 sensors in each axis.

Fig. 2. Experimental setup (left) in CT scanner environment showing (a) Aurora field generator, (b) positioning system, (c) phantom mounted on a non-metallic extension arm, (d) CT gantry and (e) CT table. - Calibration phantom (right) with two 5-DOF sensors in each axis residing 10mm from each other.



(a) (b)

Fig. 3. Comparison between relative and absolute error in different environments

A rigid transformation is found between these two sets of data. The translation vector of this transformation provides the distance between the two sensors. The same procedure is applied to other sensors and the distances between sensors for all axes are found.

Experimental setup (shown in Fig. 2) is composed of three main components: a three-axis positioning robot (Velmex Inc., Bloomfield, NY), an NDI Aurora electromagnetic tracking system (Northern Digital Inc., Waterloo, Canada) and an in-house phantom containing 6 EM sensors. The robot provides three orthogonal degrees of motion that was used to position the phantom in space. For each sampled point, the robot moved the phantom to each position, pausing at each location to allow position measurements to be collected from the Aurora for all 6 sensors. The positioning system provided the ground truth of expected positions, whereas the Aurora dataset contained the observed measurements. The Aurora position data was registered to the positioning system data using a rigid body transformation minimizing the least squares L2-norm. The absolute error is computed as the error between the EM data and the ground truth. The relative error on the other hand does not require any information from the robot measurement (ground truth). It is the difference between the EM data as measured in real-time and the calibrated geometry of the sensors.

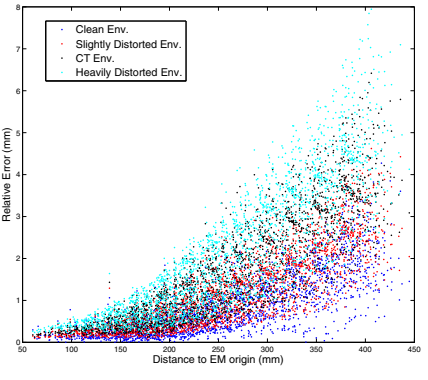


Fig. 4. Relative error in different environments

Fig. 3(a),(b) illustrate a point to point comparison between absolute and relative error for the sensors in X axis for a workspace of $30 \times 30 \times 30$cm with resolution of 2.5cm. Data were collected in two different environments; a clean

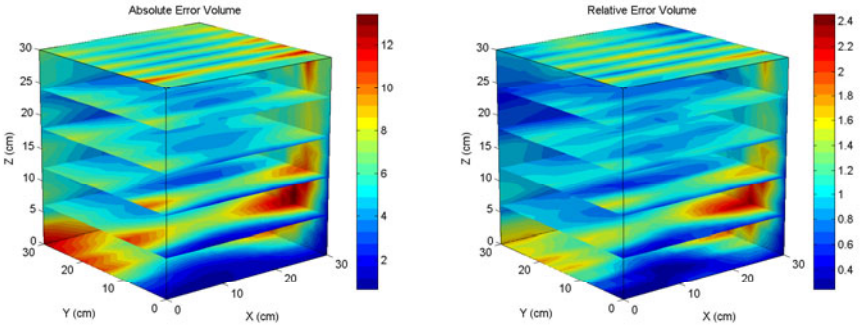
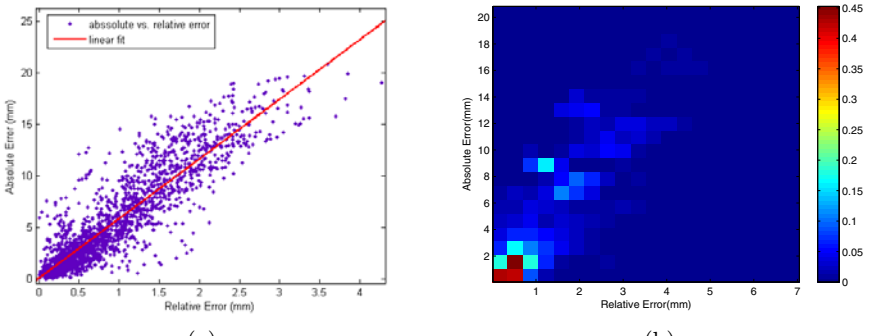


Fig. 5. Absolute (left) and relative (right) error maps in a $30 \times 30 \times 30$cm workspace of EM tracker in an interventional X-ray environment - colorbars are shown in mm

(distortion free), and a distorted (with a magnetic field next to the field generator) environments. Fig. 4 displays the relative in different environments, representing the same behavior of relative error in different environments. The data is fairly scattered and they increase sharply further they are from the origin of EM field generator. Clearly the relative error is more homogenous and shows smaller values in both environments. This is demonstrated in the absolute and relative error maps in Fig. 5, rendered from data collected in an interventional X-ray environment (Philips Allura FD20). Here, the EM field generator is parallel to the Z axis facing the XZ plane. These error volumes show similar error patterns for absolute and relative errors implying a statistical correlation between them. The correlation sustains across different environments as it could be seen from Fig. 3 and 5 where in former distortion is caused by an external magnetic field and in latter, distortion is caused by an interventional X-ray.



(a) (b)

Fig. 6. (a) Mapping between absolute and relative error in a highly distorted environment and a linear polynomial fit. (b) The probability mapping of absolute and relative error in a CT environment based on 6 sets of experiments each including more than 13000 samples.

In order to have a model as reliable as possible, we performed several experiments (with both highly distorted and minimally distorted environments), and in each experiment we collected more than 13000 sample points. As an example, Fig. 6(a) shows the mapping between absolute and relative errors for a distorted environment for X axis sensor. This figure also shows a linear fit to the data with R-square value of 0.82 which presents a good linear relation between two variables. The final histogram will be the union of all the histograms from all different environments. These environments should replicate the real situation that the EM tracking system is being used. For each specific application, the accuracy of the estimation increases by increasing the number of sampled data. Our application requires data in a CT environment. We collected 6 sets of data each including more than 13000 points. The union of all these histograms is normalized to result the probability mapping of absolute error versus relative error. The mapping is shown in Fig. 6(b). It should be noted that the region of interest in an EM tracking system is in the area of small relative errors, e.g. $RE < 1mm$, where the majority of samples occur in this region.

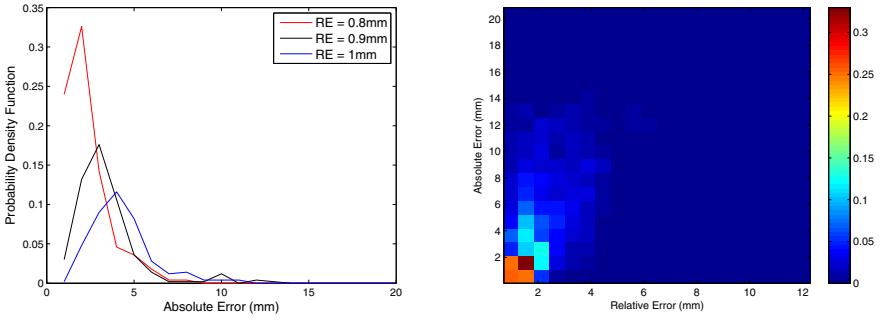
3.1 Region Based Absolute Error Estimation

In order to make the estimation of absolute error robust over a region, we collect N samples of relative error in working region and the final estimate of absolute error is the sample mean of the these N estimates (see Fig. 1). A sample variance of these N sample can serve as a confidence measure of how robust the estimation of the absolute error is in this working space. As an example, Fig. 7(a) shows the conditional probability density functions for three different values of RE ($= 1$, $= 0.8$, and $= 0.9$ mm) derived from 2-D joint histogram of Fig. 6(b). The estimated value of AE is equal to 3.89mm with a variance $\sigma^2_{AE} = 11.19$. In practice, we can collect N estimates of RE by waving the calibration phantom in the operating workspace, and estimate the statistics of AE.

4 Validation and Sensitivity Assessment

We use leave-one-out cross-validation to assess how well our probabilistic model predicts absolute errors. The joint 2-D histogram in our model is derived from the union of different environments, empirically correlating the absolute and relative errors. We leave one observation out and derive the joint histogram from the rest of environments. Then, a specific absolute error is predicted through our model and compared with the real value from the original observation. Let's consider the joint histogram derived in Fig. 6(b), and examine two areas of interest; where the absolute error is acceptable (e.g. 2mm) and where it is high (e.g. 5mm). In each iteration, we fit the model to 5 histograms and calculate the predication error for the one left out (e.g. shown in Fig. 7(b)). This is repeated 6 times for both areas of interest and the cross validation factor is calculated as:

$$CV_{2mm} = \frac{1}{6}\sum_{i=1}^{6} |\widehat{AE}_i - AE_i| = 0.46mm$$



(a) (b)

Fig. 7. (a) The probability density function of absolute error for three different readings of relative error in a CT environment. (b) The histogram for a CT environment considered for validation data.

$$CV_{5mm} = \frac{1}{6}\sum_{i=1}^{6} |\widehat{AE}_i - AE_i| = 0.89mm$$

which means a classification error of $0.46mm$ for a $2mm$ predication and $0.89mm$ for $5mm$ predication.

The sensitivity of our model is evaluated for two main parameters; the number of RE samples used for estimating AE and the number of environments used for making the histogram. For latter, we considered the histogram in Fig. 6(b) and added new environments up to 9 sets in total. For each new histogram, the estimated value of AE for the three values of RE ($= 1$, $= 0.8$, and $= 0.9$ mm) is derived. The range of estimated AEs is $4.10 - 3.44 = 0.66mm$. Our sensitivity evaluation showed that increasing the number of RE samples improves the accuracy of the AE estimate. Choosing a threshold value of 0.5mm and an area of interest of $AE = 2mm$, we increased the number of RE samples and computed the percentage of AE estimates classified truly; $|\widehat{AE} - AE| <$ threshold. Result is shown in Fig. 8. The model is specifically sensitive to smaller number of RE samples and after collecting sufficient number of samples (in our experiment 10), the estimates did not improve.

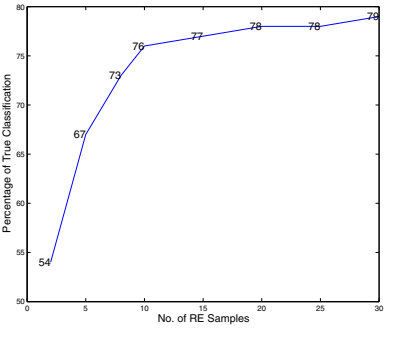


Fig. 8. True classification percentage based on the number of RE samples in a CT environment

5 Conclusion

In this paper, we applied the concept of relative error for estimating the fidelity of EM tracking field. Conventional approaches are limited to compensating errors in static environments, whereas interventional procedures often produce dynamic time varying distortions to the EM tracking field. The concept of relative error lends itself to estimating tracking error in real time by using a calibration phantom for real-time quality control of the EM field. We made a first attempt to derive an empirical statistical relationship between the relative error and absolute errors and presented a method to estimate absolute error values within a working region. In future work, we plan to use these absolute error estimates to define and visualize go/no-go regions based on thresholds provided by the operator. Intra-procedural error characterization and visual feedback in this manner can be used to alter interventional paths and/or to reposition medical instrumentation for optimal navigation accuracy. We will study the number of sensors and their orientation used together in a non-linear constellation and the size of phantom in the intra-procedural data collection procedure.

References

1. Nafis, C., Jensen, V., Jako, R.V.: Method for evaluating compatibility of commercial electromagnetic (em) microsensor tracking systems with surgical and imaging tables. In: Proc. SPIE of Medical Imaging: Visualization, Image-guided Procedures, and Modeling, vol. 9618, pp. 691–820 (2008)
2. Shen, E., Shechter, G., Kruecker, J., Stanton, D.: Effects of sensor orientation on ac electromagnetic tracking system accuracy in a ct scanner environment. In: Proc. SPIE of Medical Imaging: Visualization, Image-guided Procedures and Modeling, vol. 9618, pp. 11–23 (2008)
3. Kindratenko, V.: A survey of electromagnetic position tracker calibration techniques. J. Virtual Reality 5, 169–182 (2000)
4. Wilson, E., Yaniv, Z., Zhang, H., Nafis, C., Shen, E., Shechter, G., Wiles, A., Peters, T., Lindisch, D., Cleary, K.: A hardware and software protocol for the evaluation of electromagnetic tracker accuracy in the clinical environment: a multi-center study. In: Proc. SPIE of Medical Imaging: Visualization, Image-guided Procedures and Modeling, vol. 6509, pp. 65092T1–65092T11 (2007)
5. Frantz, D., Wiles, A., Leis, S., Kirsch, S.: Accuracy assessment protocols for electromagnetic tracking systems. J. Physics in Medicine and Biology 48, 2241–2251 (2003)
6. Fischer, G.S., Taylor, R.: Electromagnetic tracker measurement error simulation and tool design. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 73–80. Springer, Heidelberg (2005)
7. Matinfar, M., Narayanasamy, G., Guitierrez, L., Chan, R., Jain, A.: Absolute vs. relative error characterization of electromagnetic tracking accuracy. In: Proc. SPIE of Medical Imaging: Visualization, Image-guided Procedures and Modeling, vol. 7625, pp. 762524-1–762524-9 (2010)

8. Birkfellner, W., Watzinger, F., Wanschitz, F., Enislidis, G., Truppe, M., Ewers, R., Bergmann, H.: Concepts and results in the development of a hybrid tracking system for cas. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 343–351. Springer, Heidelberg (1998)
9. Feuerstein, M., Reichl, T., Vogel, J., Traub, J., Navab, N.: New approaches to online estimation of electromagnetic tracking errors for laparoscopic ultrasonography. J. of Medical Physics 13, 311–323 (2008)

A Motion Correction Algorithm for Microendoscope Video Computing in Image-Guided Intervention

Tiancheng He[1,2], Zhong Xue[1,*], Weixin Xie[2], Solomon Wong[1], Kelvin Wong[1], Miguel Valdivia y Alvarado[1], and Stephen T.C. Wong[1]

[1] The Center for Bioengineering and Informatics, The Methodist Hospital Research Institute and Department of Radiology, The Methodist Hospital, Weil Cornell Medical College, Houston, TX

[2] Intelligent Information Institute, Shenzhen University, Shenzhen, China

zxue@tmhs.org

Abstract. In multimodality image-guided intervention for cancer diagnosis, a needle with cannula is first punctured using CT or MRI -guided system to target the tumor, then microendoscopy can be performed using an optical fiber through the same cannula. With real-time optical imaging, the operator can directly determine the malignance of the tumor or perform fine needle aspiration biopsy for further diagnosis. During this operation, stable microendoscopy image series are needed to quantify the tissue properties, but they are often affected by respiratory and heart systole motion even when the interventional probe is held steadily. This paper proposes a microendoscopy motion correction (MMC) algorithm using normalized mutual information (NMI)-based registration and a nonlinear system to model the longitudinal global transformations. Cubature Kalman filter is thus used to solve the underlying longitudinal transformations, which yields more stable and robust motion estimation. After global motion correction, longitudinal deformations among the image sequences are calculated to further refine the local tissue motion. Experimental results showed that compared to global and deformable image registrations, MMC yields more accurate alignment results for both simulated and real data.

Keywords: Fluorescence microendoscopy; motion correction; normalized mutual information; Cubature Kalman filter; deformable video registration.

1 Introduction

Currently the management of small pulmonary nodules found in CT is follow-up due to the lack of specific diagnosis for biopsy. To diagnose and treat small peripheral lung cancer in an earlier stage, multimodality image-guided intervention for cancer diagnosis is a preferred technique [1]. First, the interventional needle can be punctured percutaneously with a cannula based on CT-guided intervention systems, and then optical imaging can be performed through the same cannula using fiber-optic microendoscopy. Fluorescence microendoscopy is a promising new modality for molecular imaging-guided diagnosis and is in the stage of animal studies. The

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 267–275, 2010.

FDA-approved Indocyanine Green (ICG) can be used for human, but it might not be specific to benign and malignant tumors.

Current *in vivo* microendoscopy imaging techniques include bioluminescence [2] and fluorescence techniques [3], and one task in the applications for quantitative microendoscopy image computing is that the motion in the microendoscopy image sequences needs to be corrected for better visualization and stable quantitative measures. Many methods such as image registration [4-6] were proposed for motion correction. In these methods, measures such as cross-correlation, mutual information, or object correspondences can be used to quantify the similarity between images, and geometrical transformations such as scaling, affine, and elastic deformations can be used to model the longitudinal movement. However, the longitudinal transformation parameters of image sequences obtained from registration methods are not stable because the temporal information about the image sequences has not been fully considered. To improve the longitudinal consistency, temporal smoothness constraints were used in registration of image sequences [7, 8].

In this paper, we use the normalized mutual information (NMI) to calculate the similarity between consequence frames, and at the same time, model the longitudinal transformations using a nonlinear system. The Cubature Kalman filter (CKF) algorithm [9] is used to estimate the underlying transformations. After longitudinal global alignment, the deformable transformations across the image sequences are also calculated to refine the local tissue movement by embedding the bilateral filter (BF) in the optical flow calculation [10].

Two sets of experiments were performed to evaluate the proposed MMC algorithm. First, simulated microendoscopy image sequences are used to compare quantitatively the performance of the NMI-based global with MMC. The results showed that MMC yields more accurate motion correction results. Then, we applied the algorithms to real microendoscopy image sequences. The average normalized mutual information between the baseline frame and the frame after motion correction was calculated to evaluate the performance of the motion correction. The results also confirmed better image alignment results of microendoscopy videos using MMC.

2 Method

We developed a minimally invasive multimodality image-guided (MIMIG) intervention prototype system aiming at accurately targeting the region of interest embedded in deep tissue (e.g., peripheral lung tumor) and performing in vivo diagnosis. As shown in Fig. 1, after successfully puncturing the needle inside the tumor, fiber-optic microendoscopy can be used to visualize the tumor cells with high-resolution videos (the field of view is about ~500 μm, the size of each frame is 266×264 pixel and the frame rate is 12 f/s). With proper contrast agents such as IntegrinSense (VisEn Medical Inc. Bedford, MA) targeting specific pathways of cancer cells or ICG, the needle can be locally adjusted to either perform a local molecular diagnosis or guide the fine needle aspiration biopsy to confirm cancerous tissue.

This work uses IntegrinSense as the contrast agent and the Cellvizio 660 system [11] to collect microendoscopy data for quantification of tumor and normal tissues after MIMIG intervention. We utilized the VX2 tumor model in eight White New

Zealand Rabbits in the experiments. The goal was to test the feasibility of possible cancer detection using fluorescence microendoscopy. In this paper, we present an effective motion correction method of the microendoscopy videos for better visualization and image quantification. Before motion correction, in the pre-processing step, we calculated the image similarity between consequent images of the entire video (generally last for 2-3 minutes) into shorter video clips to make sure there is no dramatic "scene" change in each clip and then focused on stable motion correction of each video clip.

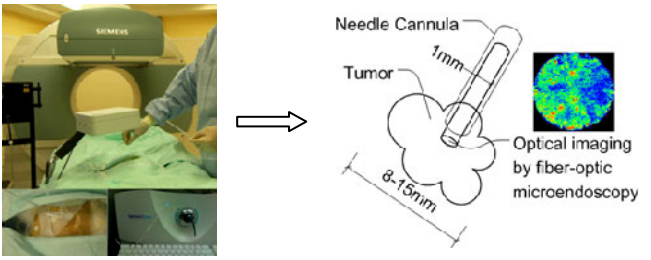


Fig. 1. Optical imaging-guided microendoscopy. After CT-guided intervention (a), a needle cannula is placed to target the tumor, and an optical fiber is inserted in the cannula, and in vivo microendoscopy is performed for high-resolution optical imaging analysis (b).

2.1 Modeling the Longitudinal Transformations Using a Nonlinear System

Given an image sequence $I=\{I_1,I_2,...,I_N\}$ with N as the number of frames, traditional registration methods calculate the global longitudinal transformations across the video by maximizing the normalized mutual information,

$$E_{NMI}(H)=\frac{1}{N-1}\sum_{t=1}^{N-1}[-NMI(H_t\circ I_t,I_{t+1})], \tag{1}$$

where $H=\{H_1,H_2,...,H_{N-1}\}$ denotes the longitudinal transformations, and both rigid and affine transformations can be fit into this model. Herein, H consists of serial translation, rotation, and scaling on actual images. $H_t\circ I_t$ is the globally transformed image I_t onto image I_{t+1}. Because there is no temporal information used in the alignment, the resultant serial alignment parameters may not be temporally stable. We propose to use a nonlinear system to model the longitudinal transformations, i.e., H can be modeled as the output of a nonlinear system as follows,

$$\begin{cases} M_t=f(M_{t-1},u_t)+n_t \\ H_t=g(M_t)+w_t \end{cases}, \tag{2}$$

where M_t is the state of the system, and u_t is the system input. n_t and w_t are independent system and measurement Gaussian noises. $f(\cdot)$ is the nonlinear system function and does not need an explicit form when the Cubature Kalman filter (CKF)

is used. The system output function $g(\cdot)$ is assumed to be an identity transformation. Also, the input signal u_t is assumed to be zero. The longitudinal transformations obtained from Eq.(1) can be regarded as an observation or output of the nonlinear system, H , and M_t is the actual transformation or the state of the system to be estimated. We hypothesize that M_t as the underlying longitudinal transformations will be more robust and accurate than H . For this purpose CKF is used to estimate M from H .

Like the Kalman filter, CKF is a minimum mean-square error (MMSE) estimator specially designed for nonlinear systems. CKF can provide a systematic solution for high-dimensional nonlinear filtering problems and is more accurate in capturing the true mean and covariance of system states. CKF converts the nonlinear system into a linear system after transferring the cubature points using the spherical-radial rule, thus it avoids calculating the Jacobians and Hesssians matrices as in Extended Kalman Filter (EKF). Moreover, CKF is more efficient than EKF for M_t estimation.

2.2 The MMC Algorithm

Combining the nonlinear system model and the NMI-based registration, the new energy function of the proposed MMC algorithm is designed as:

$$E_g(M) = \frac{1}{N-1}\sum_{t=1}^{N-1}[-NMI(H_t \circ I_t, I_{t+1}) + \alpha \left\| H_t - f(M_t) \right\|^2], \tag{3}$$

where α is the weighting factor. The first term is the same as Eq.(1), and the second term embeds the longitudinal transformation system into the image alignment procedure. H can be regarded as an observation or output of the nonlinear system for aligning the images, and M is the actual transformation to be estimated. We adopt an iterative strategy to solve M by minimizing the objective function. First, assuming M is fixed, the optimal transformations H can be calculated by maximizing NMI, and at the same time, constraining that H should be similar to M ; Then, assuming H is fixed, estimation of M_t can be achieved by using CKF. The negative NMI is used here since we minimize the energy function (thus NMI is maximized). The transformation error in the second term is defined as the distance of two the transformed points using the two given transformation matrices. Compared to the NMI-based registration in Eq. (1), the transformation matrix H is subject to a CKF filtering, and hence it can provide more stable motion correction for microendoscopy image sequences. Fig. 2 shows the flowchart of the global alignment step. From the experiments we found that the algorithm can converge in about three iterations.

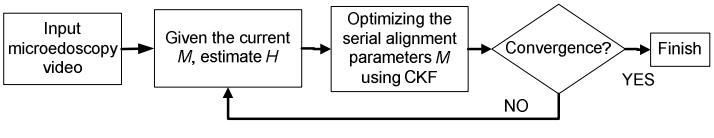


Fig. 2. The flowchart of the global registration step of the MMC algorithm

After global registration, we also applied longitudinal deformable motion correction to refine the detailed tissue movement. This step utilizes the standard deformable image registration [12]. The transformation vector v_t from the image at timepoint t to the next timepoint $t+1$ can be calculated by minimizing the following,

$$E_d(v) = \frac{1}{N-1}\sum_{t=1}^{N-1}\int_{\Omega}\left\|I_{t+1}[H_t(x)+v_t(x)] - I_t(x)\right\|^2 + \beta\left\|\nabla v_t(x)\right\|^2, \quad (4)$$

where β is the weight of the smoothness constraint. In our work, a fast 2-D implementation of this registration has been used [13]. We found that NMI is more robust for global registration since it reflects more global image features. For further local refinement using deformable registration, image intensity information is more important since 1) image intensity reflects relatively local image features, and 2) the deformable registration refinement is constrained by the result of global registration and the smoothness constraint. So both accuracy and robustness can be obtained.

3 Results

The performance of MMC is evaluated using two experiments. In the first experiment, the dataset consists of microendoscopy image sequences generated by applying simulated serial translation, rotation, and scaling on actual images captured during the image-guided intervention procedure using Cellvizio 660 system, added with Gaussian noises. The dataset of the second experiment consists of microendoscopy image sequences acquired in our image-guided intervention on lung cancer rabbit model.

3.1 Experiments Using Simulated Image Sequences

Ten simulated microendoscopy sequences using real microendoscopy frames collected from the rabbit experiments were used in this experiment. The longitudinal transformations were simulated by using sine and cosine signals,

$$p_i(t) = a_i \sin(b_i t + \phi_i) + c_i, \quad (5)$$

and $t = 1,..., N-1,\ i = 1,...,4$. a_i, b_i, and c_i are the amplitude, frequency and shifting of the transformation signals, respectively. p_1, p_2, p_3, p_4 represent the translations in x- and y-directions, rotation, and scaling. The typical amplitude for translation was set to 10 pixels, rotation angles were [-20, +20] degree, and frequency was between 0.5 and 1. The scaling range was between 0.98 and 1.02. The image sequences were generated by first transferring the reference frames using the simulated transformations and then adding spatially correlated Gaussian noises. Because only global longitudinal transformations were simulated, we compared the MMC (without deformable registration) with the NMI-based global registration. Fig. 3 shows an example of the alignment results using MMC, where the gray scale images were color-mapped according to the intensity values: red correspond to tumor cells, and the rest correspond to normal tissue and background. Fig. 3(a) shows the first frame, and five following frames were shown in the first row. In the second row, the

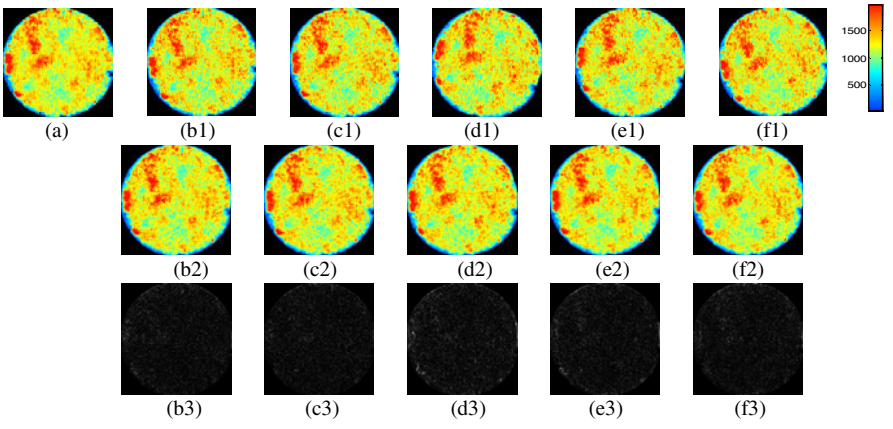


Fig. 3. Examples of results using simulated data. (a) Reference image; (b1-f1) simulated images; (b2-f2) motion corrected results using MMC; (b3-f3) the difference with the reference.

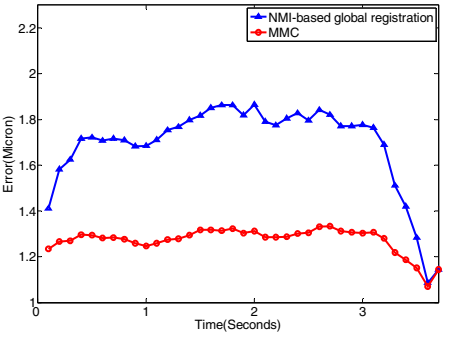


Fig. 4. Comparison of different methods: the alignment errors for each frame

corresponding aligned images were illustrated. The third row gives the difference between images. It can be seen that the difference between the reference image and the aligned image after using MMC is low.

We also calculated the average errors of longitudinal transformations between the ground truth and the alignment results. Fig. 4 shows the errors of the transformations using MMC and NMI-based global registration respectively. It can be seen that the MMC method yields more accurate estimation of the longitudinal transformations. Table 1 shows the average errors and standard deviations of the ten simulated microendoscopy videos. It can be seen that the errors of MMC are smaller than the NMI-based global registration. These results confirm that because of the use of the CKF, MMC can obtain more accurate motion correction based on our simulated image sequences. Paired t-test also showed that such accuracy improvement is statistically significant ($p<0.05$).

Table 1. The average errors and std for the ten simulated image sequence (unit in micron)

Index	NMI-based global registration	MMC (global only)
10 experiments	2.2 (0.3)	1.3 (0.1)

3.2 Microendoscopy Video Results

For validating our approach in real microendoscopy video, we used the microendoscopy videos collected from the rabbit experiments during image-guided intervention. After cutting the whole microendoscopy videos into different clips based on image similarities, we applied the MMC algorithm on 30 microendoscopy video clips. For each video clip, a frame with the smallest difference to its neighboring frames was selected as the reference frame, and all the other frames were aligned with this reference frame. Similarly we applied the full MMC algorithm, NMI-based global registration, and NMI-based global registration plus the same deformable registration method on these real image sequences. Fig. 5 shows a result using MMC. It can be seen that after motion correction the difference images is much smaller. We also calculated the NMI values between the reference frame and the aligned subsequent frames as shown in Table 2.

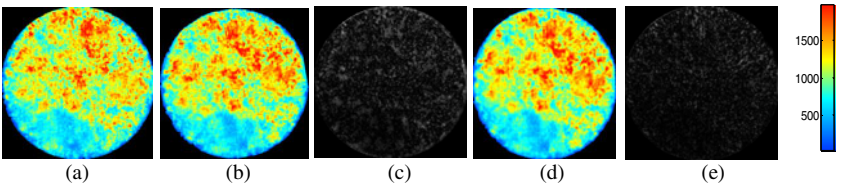


Fig. 5. Example results for real microscopy video. (a) The reference image; (b) a frame of the simulated image sequence; (c) the aligned image of (b); (d) the difference image between (b) and (a); (e) the difference image between (c) and (a).

Using the proposed MMC the accuracy of image alignment is improved for all the 30 image sequences studied. If we compare the results between NMI-based global registration and NMI-based global registration + deformable registration, we can see the improvements on image alignment by using additional deformable registration step. If the MMC and the method using NMI-based global registration + deformable registration are compared, we can see the improvements because of use of CKF in modeling the longitudinal transformations. In average, the NMI values are improved for 12.7% by comparing MMC with the NMI-based global registration. Our future work will be implementing the method in the image-guided system for visualization of the microendoscopy video and for quantitative analysis to establish the criteria for tumor detection.

Table 2. The average NMI values in each video

Index	(a)NMI-based global registration	(b)NMI-based global registration + deformable registration	(c)MMC	Improvement between (b) and (c) (%)
30 experiments	0.58 (0.07)	0.63 (0.08)	0.70 (0.08)	12.7% (4.1%)

4 Conclusion

Microendoscopy imaging is a potential powerful tool used in image-guided intervention for better guidance in lung cancer diagnosis and treatment procedures. In order to generate better visualization of the microendoscopy videos or to quantitatively calculate stable features to distinguish tumor tissues from normal tissues, an efficient and stable video alignment is needed. This paper proposes a longitudinal video motion correction method by applying a nonlinear system to enforce the longitudinal stability of global transformations and using a global and deformable registration to find the detailed tissue movement. More accurate alignment results can be obtained using both simulated and real microendoscopy image sequences. In the future, we will apply MMC to our molecular image-guided intervention system for peripheral lung cancer and study quantitative methods to characterize and differentiate different tissue types of the lung. MMC may also be extended to other object tracking applications in computer vision.

References

[1] He, T.C., Xue, Z.: A Minimally Invasive Multimodality Image-Guided (Mimig) Molecular Imaging System for Peripheral Lung Cancer Intervention and Diagnosis. In: Proceeding of Information Processing in Computer-Assisted Intervention (2010)

[2] Gheysens, O., Mottaghy, F.M.: Method of Bioluminescence Imaging for Molecular Imaging of Physiological and Pathological Processes. Methods 48(2), 139–145 (2009)

[3] Tian, J., Bai, J., Yan, X.P., Bao, S., Li, Y., Liang, W., Yang, X.: Multimodality Molecular Imaging. IEEE Eng. Med. Biol. Mag. 27(5), 48–57 (2008)

[4] Jenkinson, M., Bannister, P., Brady, M., Smith, S.: Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. Neuroimage 17(2), 825–841 (2002)

[5] Vercauteren, T., Perchant, A., Malandain, G., Pennec, X., Ayache, N.: Robust Mosaicing with Correction of Motion Distortions and Tissue Deformations for in Vivo Fibered Microscopy. Medical Image Analysis 10(5), 673–692 (2006)

[6] King, A.P., Jansen, C., Rhode, K.S., Caulfield, D., Razavi, R.S., Penney, G.P.: Respiratory Motion Correction for Image-Guided Cardiac Interventions Using 3-D Echocardiography. Medical Image Analysis 14(1), 21–29 (2010)

[7] Shen, D.G., Sundar, H., Xue, Z., Fan, Y., Litt, H.: Consistent Estimation of Cardiac Motions by 4d Image Registration. Medical Image Computing and Computer-Assisted Intervention - Miccai 2005, Pt 2. 3750, 902–910 (2005)

[8] Castillo, E., Castillo, R., Martinez, J., Shenoy, M., Guerrero, T.: Four-Dimensional Deformable Image Registration Using Trajectory Modeling. Phys. Med. Biol. 55(1), 305–327 (2010)

[9] Arasaratnam, I., Haykin, S.: Cubature Kalman Filters. IEEE T. Automat Contr. 54(6), 1254–1269 (2009)

[10] Xiao, J.J., Cheng, H., Sawhney, H., Rao, C., Isnardi, M.: Bilateral Filtering-Based Optical Flow Estimation with Occlusion Detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 211–224. Springer, Heidelberg (2006)

[11] Sonn, G.A., Mach, K.E., Jensen, K., Hsiung, P.L., Jones, S.N., Contag, C.H., Wang, T.D., Liao, J.C.: Fibered Confocal Microscopy of Bladder Tumors: An Ex Vivo Study. Journal of Endourology 23(2), 197–201 (2009)
[12] Shen, J., Matuszewski, B.J., Shark, L.: Deformable Image Registration. In: Proceeding of IEEE ICIP, pp. 1112–1115 (2005)
[13] Wei, L., Shen, D.: Effect of Hierarchical Deformable Motion Compensation on Image Enhancement for Dsa Acquired Via C-Arm. In: Proceeding of the 20th Annual IS&T/SPIE Symposium on Electronic Imaging, pp. 27–31 (2008)

Least-Incision Transformable End-Effector Mechanism for Forceps for Endoscopic Surgery

Hiroaki Nakaji and Ryoichi Nakamura

Department of Medical System, Division of Artificial Systems Science,
Graduate School of Engineering, Chiba University
1-33 Yayoi-cho, Inage-ku, Chiba-shi 263-8522, Japan
hiroaki.n@graduate.chiba-u.jp

Abstract. The demand for thinner instruments for endoscopic surgery has been increasing because thinner instruments minimize invasiveness and increase the applicability of endoscopy. However, in thinner instruments, the end effectors are smaller, which limits instrument functionality. We have developed a new pair of forceps using a least-incision transformable end-effector (LITE) mechanism that transforms its end effectors by increasing its size within the body cavity. In our experiments, the grasping force was measured to be greater than 5.3 N. Five non-specialists and three surgeons performed endoscopy to evaluate the end effector's transformation and removal times. The average transformation and removal times were 108 ± 37 s and 65 ± 23 s and 86 ± 37 s and 61 ± 27 s, respectively, for the non-specialists and surgeons, respectively. An in vivo experiment was also conducted on a pig using the LITE forceps. Our mechanism is extremely useful for performing least-invasive endoscopy surgeries.

Keywords*:* Endoscopic surgery, Transformable end effector, Forceps, Least-invasive.

1 Introduction

Today, the demand for thinner instruments for endoscopic surgery, including fetal and single-port surgeries, has been increasing since thinner instruments reduce the invasiveness of endoscopic procedures, improve the applicability of endoscopy, and promote the development of new devices for new procedures [1][2].

However, the thinner the instruments, the smaller their end effectors, which severely limits their functionality. Therefore, it is important to develop instruments that become sophisticated within body cavities. A retractor is a medical instrument used in endoscopic surgeries. Once inserted into the body cavity, it can open in a fan-like manner to displace organs, but it cannot perform complex movements. To achieve finer and more selective control, "assemblable hands" that can be assembled inside the abdominal cavity and become large instruments have been developed [3]–[5]. However, this is sometimes risky since parts of the device might detach and become lost.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 276–285, 2010.

In this study, we have developed a novel design for a pair forceps—a least-incision transformable end-effector (LITE) mechanism is used to transform the instrument into a larger end effector inside the body. The LITE mechanism is initially shaped like a thin rod before it is inserted. This enables surgeons to use forceps whose end effectors are larger than the port size. As a result, there is no risk of loss of device parts because all its components are connected. Furthermore, we have also implemented our LITE mechanism in a pair of forceps and evaluated its efficacy through experiments.

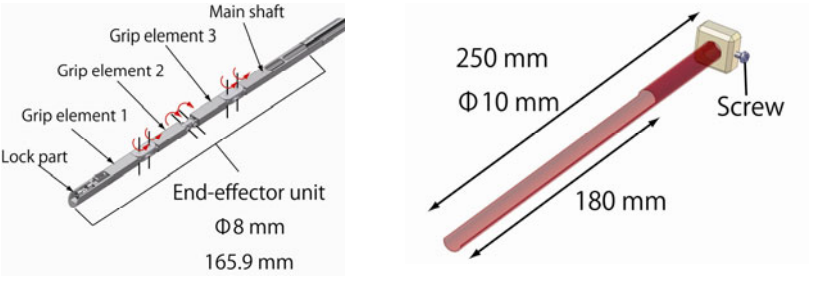
2 Methods

2.1 Requirements

Our aim is to develop an end effector that transforms into a pair of forceps for applications to surgical operations. Hence, the form of the LITE mechanism after the transformation has to be the grasper form, and its width in the insertion axis has to be in the perpendicular direction. In a first trial, the diameter of the end effector is less than 10 mm. Subsequently, we aim to develop an end effector with the diameter less than 5 mm.

2.2 LITE Mechanism

Fig. 1 shows the structure of the LITE mechanism. The mechanism has six pin joints and one degree of rotational freedom at each pin joint (Fig. 1(a)). The pipe is used to support the transformation and removal procedures of the mechanism (Fig. 1(b)). A part of the pipe has a semicircular cross-section that allows the LITE mechanism to bend at a specific joint, and it remains straight otherwise.



(a) (b)

Fig. 1. (a) Transformable end effector for endoscopic surgery and (b) auxiliary pipe

2.3 Transformation and Fixation Mechanisms

The transformation process is outlined in Fig. 2. The LITE mechanism transforms by bending in two stages, each of which is controlled by a wire connected to the tip and

tensioned by the user from outside the abdominal cavity. The transformation process consists of the following steps:

Step 1. Insert the LITE mechanism into the body cavity using an auxiliary pipe.
Step 2. Pull wire 1. The LITE mechanism rotates about axes 1 and 2 (first transformation).
Step 3. Remove the auxiliary pipe.
Step 4. Pull wire 2. The LITE mechanism rotates about axes 3 and 4 (second transformation).
Step 5. The transformation is completed.
Step 6. The device can be held and operated via the slider rod.

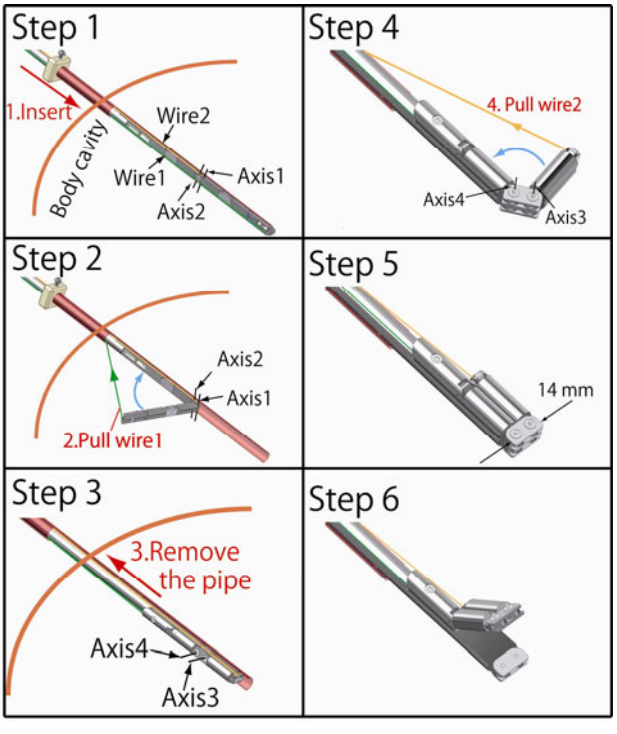


Fig. 2. Proposed transformation procedure

It is necessary to fix the grip elements onto the main shaft after the first and second transformations. Fig. 3 shows the position of the lock element and Fig. 4 shows the lock mechanism used for the first transformation. The fixation parts first close in step 3; this is the "unlock" mode. The lock part is fastened to the main shaft by the fixation parts that open sideways when the rod is inserted; this is the "lock" mode. A spring is assembled inside the fixation parts that can then be closed by the force of the spring displacing the rod (back to unlock mode). This mechanism makes it possible to disassemble and remove the device from the body cavity.

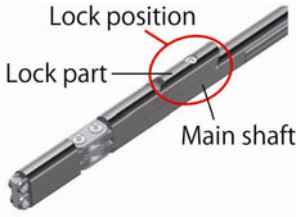


Fig. 3. Lock position

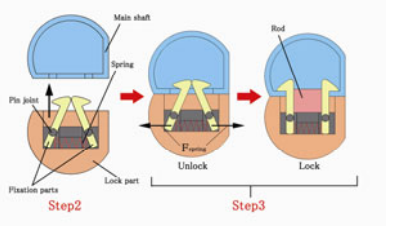


Fig. 4. Lock mechanism

Fig. 5 shows the structure and position of wire 2. The transformation shown here is achieved by the user pulling wire 2 (Step 4). Therefore, it is possible to fasten the grip element 3 to the main shaft by the tension force of wire 2 that is fixed outside the abdominal cavity. The tension of wire 2 is sufficient for performing the "lock" operation of the second transformation because this position is locked mechanically when the gripper opens.

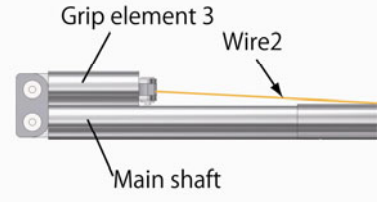


Fig. 5. Position of wire 2

2.4 Grasping Mechanism

The gripper of the LITE mechanism increases in size after the transformation. Hence, to achieve a sufficiently strong grip, a large generative force is required. Therefore, the grasping mechanism is implemented with a slider link, as shown in Fig. 6. The link is inserted inside a notch in the slider rod during the first transformation. The grasper is then driven by pushing and pulling the slider rod.

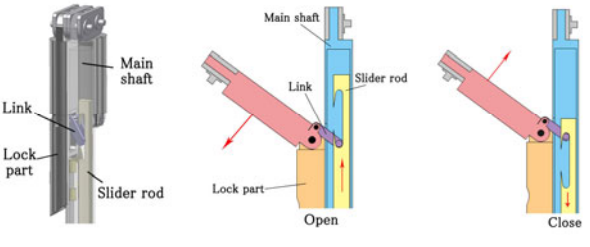


Fig. 6. Grasping mechanism

2.5 Removal Procedure

Fig. 7 outlines the process for removing the device from the body cavity. This process is an exact reversal of the previously performed transformation. The LITE mechanism is straightened by pushing on the auxiliary pipe and can then be removed.

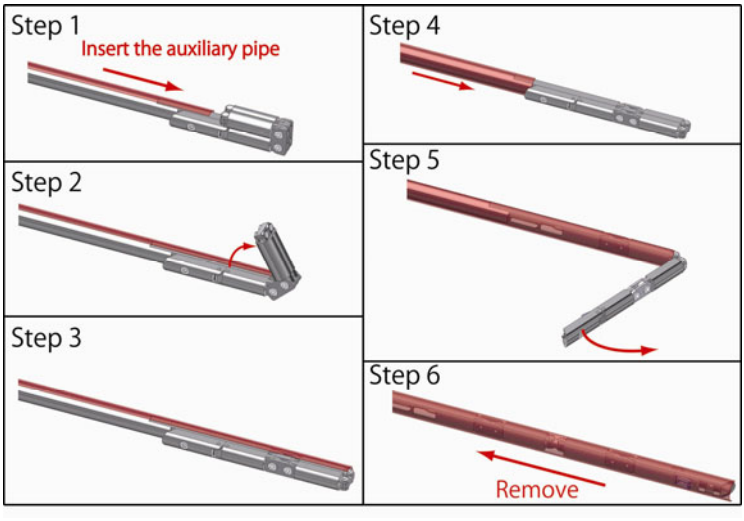
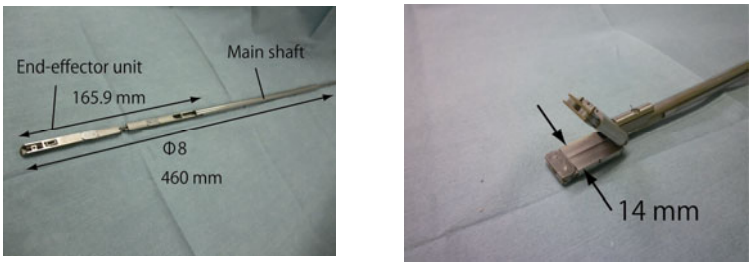


Fig. 7. Removal procedure

3 Prototype

Fig. 8 shows the device prototype; it has a diameter of 8 mm and a length of 460 mm before the transformation. The width of the gripper is 14 mm.



(a) Before transformation (b) After transformation

Fig. 8. Photographs of device prototype

4 Experiments

4.1 Measurement of Grip Force

We measured the grip force exerted by the prototype by using the experimental system shown in Fig. 9. A weight, providing an input force F_{in} between 0 N and 50 N, was attached to the end of the slider rod, and the resulting grip force was measured five times using a digital force gauge (FGPX-5, Nidec-Shimpo Corporation, Kyoto, Japan). We also calculated the average value for each F_{in}. For comparison, we also measured the grip force achieved with two traditional graspers—a 5-mm-diameter grasper with a ratchet handle and a 10-mm-diameter grasper (Babrock with ratchet handle, Ethicon Endo-Surgery, Inc.). The interfaces of these traditional graspers were removed and a weight was attached directly to the gripper rod. Fig. 10 shows the results of the grip force measurements. The grip force exerted by the LITE mechanism closely follows that exerted by the 5-mm-diameter grasper. The grip force with an input force of 50 N was 5.3 N. The maximum standard deviation (SD) of the measured grip force of the LITE mechanism was 0.11 N.

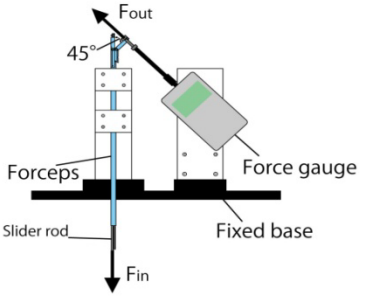


Fig. 9. Experimental system

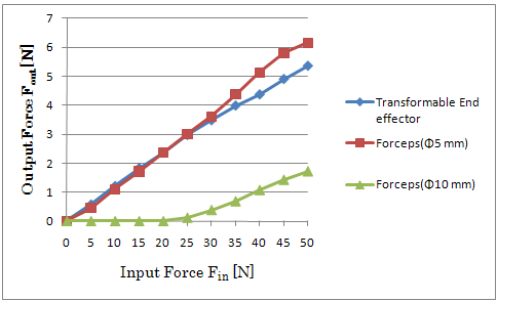


Fig. 10. Experimental results

4.2 Measurement of Transformation and Removal Times

We measured the time required for transformation and removal in a simulated endoscopic surgery environment using an endoscopic surgery trainer (MATT Trainer, Limbs & Things).

The transformation and removal of the LITE mechanism was performed by eight persons, five of whom were non-specialists (students) and three were surgeons with experience in laparoscopic surgery. Each person performed the operation thrice; they controlled the LITE mechanism with both hands without assistance and without using additional forceps. Fig. 11 shows photographs of the transformation experiments.

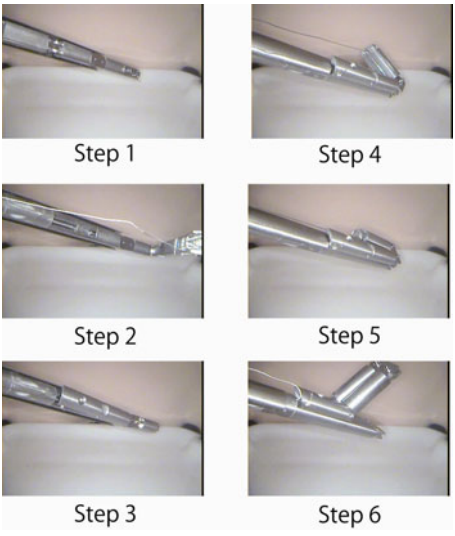


Fig. 11. Transformation experiments

Table 1 lists the measurement results for the non-surgeons; the average transformation time was 108 ± 37 s, and the average removal time was 65 ± 23 s. Table 2 lists the corresponding results for the surgeons; here, the average transformation time was 86 ± 37 s, and the average removal time was 61 ± 27 s. The transformation and removal times of the non-surgeons and professional surgeons are different, but, in general, both groups completed the procedures within 1~2 min.

Table 1. Time taken for transformation and removal (non-surgeons)

Test subject	Transformation Sequence [s]				Removal Sequence [s]			
	Max.	Min.	Av	Sd.	Max.	Min.	Av	Sd.
A	92	76	85	8	65	28	49	19
B	202	75	127	67	76	59	68	9
C	161	87	128	38	72	61	66	6
D	134	109	122	13	92	78	87	8
E	94	65	77	15	86	48	62	21
Total	202	65	108	38	92	28	66	17

Table 2. Time taken for transformation and removal (surgeons with experience in laparoscopic surgery)

Test subject	Transformation Sequence [s]				Removal Sequence [s]			
	Max.	Min.	Av	Sd.	Max.	Min.	Av	Sd.
F	152	63	97	48	84	40	65	23
G	150	59	91	51	102	48	69	29
H	79	60	70	10	89	27	49	35
Total	152	59	86	37	102	27	61	27

4.3 In Vivo Experiment

An in vivo experiment was conducted on a pig under simulated laparoscopic surgery conditions, as shown in Fig. 12.

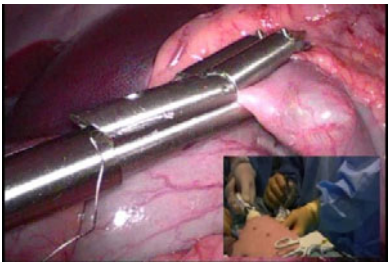


Fig. 12. In vivo experiment

This experiment confirms that the transformation and removal of the LITE mechanism in the abdominal cavity is possible. However, its tip impacted the abdominal wall and internal organs, but there was no damage. Although the smoothness of the gripper surface made it slippery, it was possible to grasp sections of the small intestine. We weren't able to confirm that space of operative field was worse because of using LITE mechanism. Hence, there was no problem with a pneumoperitoneum gas leak.

5 Discussion

5.1 Grip Force

The pinch force necessary to avoid slippage is more than 3.3 N [6]. Therefore, the result that a grip force of 5.3 N can be achieved with an input force of 50 N shows that the LITE mechanism can grip without slippage. Moreover, it can be used as a grasper because its grip force equals that of a 5 mm-diameter grasper. On the other hand, the grip force of the 10 mm-diameter grasper is below the expected value. Hence, it is difficult to pull the rod used for driving the gripper with a low input force because of the strong friction in the internal mechanism. To investigate the maximum grip force, we performed an additional experiment—we measured the grip force of the 10 mm-diameter grasper by gripping the handle with the hand. The maximum grip force was more than 30 N. We also measured the grip force of the LITE mechanism that was attached to a simplified interface in a similar manner. The grip force of the LITE mechanism was more than 30 N. Thus, we can conclude the LITE mechanism of power transmission is adequate.

5.2 Transformation and Removal Times

The average transformation time achieved by the professional surgeons was 22 s faster than that by non-surgeons. This is because the non-surgeons are not accustomed to operating surgical instruments while watching an endoscopic monitor.

The professional surgeons who participated in this experiment were of the opinion that the transformation time must be less than a reasonable threshold for the LITE mechanism to be useful as a grasper. Therefore, it is necessary to improve the transformation mechanism in order to shorten the transformation time. We used an auxiliary pipe to support the transformation, but this made the procedure time-consuming because it is difficult to simultaneously operate the LITE mechanism and the auxiliary pipe. Hence, another approach that does not use an auxiliary pipe must be found. The professional surgeons and non-surgeons were able to remove the LITE mechanism within approximately 1 min, with no significant differences between these two groups. This procedure is faster than the transformation procedure because it is considerably simpler.

5.3 In Vivo Experiment

A disadvantage of performing the transformation inside the abdominal cavity is that the tip of the LITE mechanism, being long, impacts the abdominal wall and internal organs. Hence, it is necessary to improve the mechanism by shortening the bending radius to avoid causing any damage. The LITE mechanism has three bending joints, and it bends at the middle joint during the first transformation. Therefore, the bending radius can be shortened by bending the mechanism at the joint closest to the tip. Furthermore, the fixations and link used for grasping collide with the other components during the transformation when the current locking and grasping mechanisms are used. Therefore, the locking and grasping mechanisms have to be improved.

However, we confirm that the mechanism is able to grasp and lift the small intestine successfully. According to the surgeons' advice, the gripper has a sufficiently large area for grasping larger organs. The solution to the problem of slippage could be to increase friction at the surface by mechanical processing.

5.4 Future Work

We have adopted a wire-driven mechanism for performing the transformation, but other mechanisms using springs and gears also exist. In the future, we will consider a safer mechanism that realizes the transformation instantly. We also aim to reduce the diameter of the LITE mechanism to 5 mm. In fetoscopic surgery, to make the invasion of the delicate uterine wall as minimal as possible, the diameter of the instruments must be minimized [2]. Moreover, thin instruments minimize invasions in other endoscopic surgeries as well. Our ultimate aim is to develop a LITE mechanism with a diameter of 3 mm. In this study, the transformation mechanism was implemented to develop a pair of forceps. We aim to apply this mechanism to other types of devices in the future.

6 Conclusion

We have developed a new pair of new forceps using the LITE mechanism in order to meet the requirements of thinner endoscopic instruments. Our experimental results

show that the grasping force achieved was greater than 5.3 N, while the average time taken by surgeons for the transformation of the LITE mechanism was 86 s. We also evaluated the basic performance of the device by conducting an in vivo experiment on a pig. Our results confirm that the LITE mechanism can effectively enhance the functionality of forceps.

Acknowledgement

This study was supported in part by a Grant-in-Aid for Young Scientists (B) (MEXT, Japan), Special Coordination Funds for Promoting Science and Technology of MEXT, and a research grant from CASIO Science Promotion foundation. We would like to thank Prof. Tatsuo Igarashi, Department of Medical Systems, Division of Artificial Systems Science, Graduate School of Engineering, Chiba University; Prof. Hideki Hayashi and Prof. Haruhumi Makino, Research Center for Frontier Medical Engineering, Chiba University; and Prof. Yukio Naya, Teikyo University Chiba Medical Center, for their assistance in the evaluation experiment.

References

1. Ikuta, K., Sasaki, K., Yamamoto, K., Shimada, T.: Remote Microsurgery System for Deep and Narrow Space – Development of New Surgical Procedure and Micro-robotic Tool. In: Dohi, T., Kikinis, R. (eds.) MICCAI 2002. LNCS, vol. 2488, pp. 163–172. Springer, Heidelberg (2002)
2. Yamashita, H., Matsumiya, K., Masamune, K., Liao, H., Chiba, T., Dohi, T.: Miniature Bending Manipulator for Fetoscopic Intrauterine Laser Therapy to Treat Twin-to-Twin Transfusion Syndrome. Surgical Endoscopy 22, 430–435 (2008)
3. Takayama, T., Omata, T., Futami, T., Akamatsu, H., Ohya, T.: Detachable-Fingered Hands for Manipulation of Large Internal Organs. In: 2007 IEEE International Conference on Robotics and Automation, Roma, pp. 244–249 (2007)
4. Ohshima, R., Takayama, T., Omata, T., Ohya, T.: Assemblable Three Fingered Five-DOF Hand for Laparoscopic Surgery. In: 2008 IEEE International Conference on Robotics and Automation Pasadena, CA, pp. 3896–3901 (2008)
5. Osaki, M., Takayama, T., Omata, T., Ohya, T., Kojima, K., Takase, K., Tanaka, N.: Single-Trocar Assemblable Retractor-Hand for Laparoscopic Surgery. In: 2009 IEEE International Conference on Robotics and Automation, Kobe (2009)
6. Heijinsdijkp, E.A.M., Pasdeloup, A., Dankelman, J., Gouma, D.J.: The Optimal Mechanical Efficiency of Laparoscopic Forceps. Surgical Endoscopy 18, 1766–1770 (2004)

Real-Time Organ Tracking in Ultrasound Imaging Using Active Contours and Conditional Density Propagation

Xiaohui Zhang[1], Matthias Günther[2], and André Bongers[2,⋆]

[1] Klinikum Mannheim, University of Heidelberg, Mannheim, Germany
[2] Medical Imaging Research Institute, mediri GmbH, Heidelberg, Germany
a.bongers@mediri.com

Abstract. Ultrasound tracking of organs or target volumes is a promising means to correct the displacement caused by respiration and errors from repositioning in medical applications e.g. in radiation therapy. However, one major problem of ultrasound images is their inherent low contrast and clutter which often makes standard algorithms instable for this purpose. In this work we present the adaption and application of a probabilistic tracking approach based on conditional density propagation (condensation) for real-time tracking on ultrasound images. This approach promises to facilitate robust real-time tracking with 5 degrees of freedom (translation and scaling in x-/y- direction, rotation) of anatomic structures on noisy and low contrast ultrasound images. The real-time performance and precision of the algorithm are investigated with ultrasound data from the liver. The tracking results of the algorithm are compared with results obtained from image registration. It is shown that this algorithm is real-time capable with processing time less than 5 ms per frame and robust on low contrast target structures with a precision below 1.6 mm in translation. Compared with an independent image co-registration method, this method leads to a superior displacement correction in pre-delinated target structures.

1 Introduction

In modern radiotherapy, the procedures of delivering prescribed dose are always affected by the inevitable organ motion though machines have precision in millimetre range. Real-time tracking of the organ motion can help to improve the precision by simultaneously and synchronically adjusting the target position. Due to its real-time, portable and non-invasive nature, ultrasound imaging suggests itself a good choice for tracking. Although substantial work has been done to facilitate ultrasound tracking of organ motion the problem of robust motion detection is still partly unsolved. The substantial clutter and low contrast forms a

⋆ Additionally, we would like to thank Prof. Dr. Frank Lohr, Dr. Judit Boda-Heggemann and Dr. Martin Griebe from the University of Heidelberg for the support with the acquisition of the ultrasound image data.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 286–294, 2010.

major obstacle in motion control in ultrasound images. For non-real time applications feasible solutions have been proposed to e.g. retrospectively segment organ boundaries on ultrasound series in echocardiography [1,2,3]. Also, the tracking of slowly moving surgical instruments, which generally exhibit high contrast and a known shape, has been successfully pursued with ultrasound images [4]. However, robust (i.e. noise insensitive) real-time tracking of tissue regions still remains a challenge. Although some promising methods have been suggested to track resparitory motion in abdominal ultrasound images, these methods are generally computationally costly and require the incorporation of other imaging modalities [5] or additional markers [6].To overcome these obstacles we used an active contour algorithm that possesses beneficial properties to track noisy, low contrast ultrasound image structures. The algorithm relies on a bayesian probabilistic approach with conditional density propagation (condensation) which was first proposed by Isard and Blake [7,8]. The statistical nature of this approach makes the process stable against image noise and hence suits ultrasound image sequences well. This process was proved to be fast enough to operate and control the motion in real-time [9].

In this work we implemented the 2D real-time ultrasound tracking and investigated the motion tracking behaviour in-vivo within the liver. To facilitate ultrasound tracking with this algorithm, a dedicated measurement and weight calculation procedure was introduced. The precision of extracted trajectories was determinated by inverstigating the variances between different real-time runs of the algorithm. Besides, this contour tracking algorithm have been compared to an independent (non-real-time capable) co-registration of the ultrasound frames.

2 Methods and Meterials

2.1 Contour Tracking Algorithm

To track motion with cluttered and noisy ultrasound images an algorithm for contour tracking using a stochastic random sampling procedure (condensation algorithm) [8] was adopted in this work and a dedicated measurement and weight calculation procedure was introduced to meet the special requirements for ultrasound tracking. As a first step, a contour on a target structure is delineated manually by marking several points in one frame of the ultrasound data. In this delineation procedure points expected to be bright are marked with a bright point marker and points expected to be dark with a dark point marker. In this way a base contour with M_b bright points M_d dark points(M_b+M_d=M) is generated. To represent the movement of the base contour, a contour state s that consists of the transform parameters in 5 degrees of freedom(x and y translation, x and y scaling and rotation) is used. The tracking procedure takes a Bayesian approach and starts with the initialization of a N size contour state set $\{s_0^{(n)}, n = 1, ..., N\}$ with a gaussian number generator (in this work N was set as 1000 to enable a real-time tracking). In subsequent procedures a probability

distribution of current state set $\{s_k^{(n)}, n = 1, ..., N\}$ for current ultrasound image is calculated and a predicting state set is derived in the following steps:

1. Observation on the ultrasound data: For each sample an observation Z_k is derived by measuring the gray values I_b of bright markers and I_d of dark markers in the ultrasound frame k . The observation density $p_z(s_k^{(n)})$ for a certain sample $s_k^{(n)}, n \in \{1, ..., N\}$ is generated in the following way:

$$p_z(s_k^{(n)}) = p(Z_k|X_k = s_k^{(n)}) \propto (\frac{1}{M_b} \cdot \sum_{i=1}^{M_b}((I_b)_i) - \frac{1}{M_d} \cdot \sum_{i=1}^{M_d}((I_d)_i))^2 \quad (1)$$

Then a weight set $\{\pi_k^{(n)}, n = 1, ..., N\}$ which represents the probability of respective state set is generated as follows:

$$\pi_k^{(n)} = \frac{p_z(s_k^{(n)})}{\sum_{i=1}^{N}(p_z(s^i)}, n \in \{1, ..., N\}. \quad (2)$$

2. Sample selection: As a preparation for the next step, samples in the prior set $\{s_k^{(n)}, n = 1, ..., N\}$ with high weight are selected preferentially to generate a new N-size selected set $\{s'^{(n)}_k, n = 1, ..., N\}$, that is to say, the propagation posibility of a certain sample $s_k^{(n)}, n \in \{1, ..., N\}$ is proportional to the respective weight $\pi_k^{(n)}$.

3. Propagation: The selected set $\{s'^{(n)}_k, n = 1, ..., N\}$ then serves as the base for the following autoregressive diffusion process to generate a posterior set $\{s_{k+1}^{(n)}, n = 1, ..., N\}$:

$$s_{k+1}^{(n)} = \overline{s_k} + A \cdot (s'^{(n)}_k - \overline{s_k}) + \sigma \cdot G(0, 1), n \in \{1, ..., N\} \quad (3)$$

and

$$\overline{s_k} = \sum_{n=1}^{N} \pi_k^{(n)} \cdot s_k^{(n)}, \quad (4)$$

where $G(0, 1)$ is a Gaussian diffusion term with a width σ adjusted according to the tracking problem, while A is the factor defining the process scaling which describes the weight of the propagated sample relative to the stochastic diffusion process in the model(in this work A=0.4 was set).

The three steps above iterate themselves to complete a tracking of the target structures over all the frames in a ultrasound data. To extract a propagated contour a consensus contour is calculated as a weighted mean of all samples in the current sample set. The mean state parameters $\overline{s}$ serve to generate movement parameters for the control of a medical device.

2.2 Materials

In this work, the algorithm was implemented in C++ on a PC with 2 Pentium XEON 3.2MHz processors and 1GB RAM. The software makes use of DirectX

9 to stream ultrasound data and can be used with on-line ultrasound imaging or off-line with stream playback from a hard-drive. The algorithm performance was tested in-vivo with respect to real-time tracking of respiratory motion. The ultrasound data of the liver with approximate sagittal view were acquired in a volunteer during around 10 deep breathing cycles with an ultrasound system (Echo Blaster 128 1-Z)from Telemed Inc. (LT). For image generation a curved linear array transducer (Vermont Inc., 3.5. MHz, 128 lines, FoV=220mm) was used. The data were acquired in real-time with a frame rate of 25Hz and subsequently stored to a hard drive. The algorithm was run on the data by playing them back with the same frame rate as acquired.

3 Results

To demonstrate the feasibility of robust tracking, the proposed method was evaluated on liver ultrasound data. The parameters in the algorithm were set as follows: sample number=1000, σ(x-translation)=σ(y-translation) = 0.01, σ(x-scaling) =σ(y-scaling)= 0.005, σ(rotation) = 0.01, and the base contour was generated by manually picking up 28 contour support points (13 bright and 15 dark points) on the diaphragm and vena portae in the first frame of the ultrasound series. During the tracking, displacement parameters in 5 degrees of freedom for each ultrasound frame were recorded. An exemplary raw data frame with corresponding contour delineation is shown in Figure 1. The algorithm tracked the delineated features in real-time with a average computing time of 3.4 ms for each frame. Visual inspection of the tracking contour shows that the tracking contour follows the target structures very closely.

3.1 Algortithm Precision

To obtain a measure for the reproducibility of our stochastic procedure, several real-time runs of the algorithm on the same liver dataset were compared. In

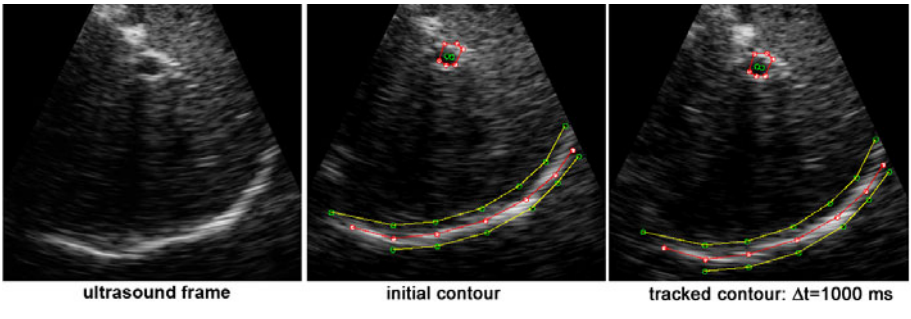


Fig. 1. Left: The field of interest in a ultrasound frame. Middle: The initial contour(28 points) overlayed on the first frame(Δt=0 ms) of the liver ultrasound data. Right: The tracking contour from our real-time algorithm overlayed on a later frame during deep breathing (Δt=1000 ms).

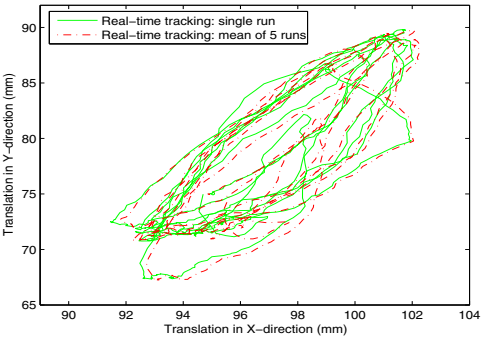


Fig. 2. x-y Trajectory of the contour with a single real-time tracking and the mean of 5 runs on the liver ultrasound data(36s, 909 frames) during deep breathing with 28 contour points(13 bright and 15 dark points). Tracking parameters were set as followed: sample number=1000, 5 degrees of freedom, σ(x-translation)=σ(y-translation) = 0.01, σ(x-scaling) =σ(x-scaling)= 0.005, σ(rotation) = 0.01.

Table 1. Standard deviation of 5 real-time runs in all the degrees of freedom with contour tracking(1000 samples, 28 contour points)

Standard deviation	Translation -X[mm]	Translation -Y[mm]	Scaling -X [100%]	Scaling -Y [100%]	Rotation [rad]	Computing time [ms]
Average	0.40	0.17	0.46	0.21	0.007	0.025
Minimum	0.071	0.027	0.082	0.036	0.002	0.001
Maximal	1.54	0.66	1.24	1.27	0.027	1.855

Figure 2 a typical real-time x-y trajectory of the contour (approx. 10 breathing cycles) is shown together with the mean over 5 independent runs of the experiment. From the figure, it can be seen that the real-time trajectory generally follows the mean trajectory within a distance of less than 1 mm. As a general measure for algorithm reproducibility the standard deviation for 5 independent real-time tracking runs on the same dataset was calculated. The maximal, minimum and average values of standard deviation in the time course are shown in Table 1, from which it can be seen that for typical situations the displacements of the live structures can be determined with a precision well below 1.6mm with the computing time for each frame less than 5ms.

3.2 Algorithm Comparison with Image Co-registration

As an independent measure of tracking quality, the results from the tracking procedure were then compared with an independent frame-by-frame image co-registration. Due to the relative poor quality of ultrasound images, all the frames of the ultrasound data were prepocessed with a mean filter of 9-by-9 neighborhood. Then co-registration was implemented using the ITK library, in which *Affine Transform,Linear Interpolator, Mean Squares Metric* and *Regular Step*

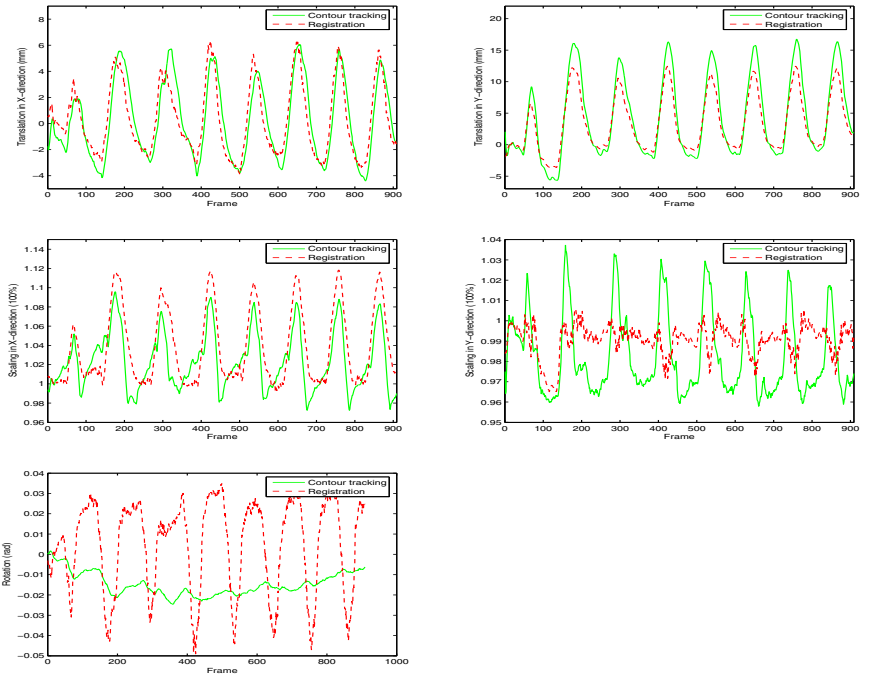


Fig. 3. Time courses of the 5 transform parameters(x-/y- translation, x-/y- scaling and rotation) for the liver movement during 8 breathing cycles resulting from the contour tracking algorithm and the co-registration method

Gradient Descent Optimizer are used [10]. Frame 15 in the liver data was defined as the base frame and then registered to all the other frames one-by-one to get the transform parameters. In this way, a continual spatial transform parameter set was generated as a contrast of the contour tracking algorithm.

Figure 3 shows the results of the 5 transform parameters with the contour tracking algorithm and co-registration method, in which frame 15 was adopted as a reference frame for both results. As we can see, the co-registration and contour tracking algorithm yielded rather similar time-courses for x-/y- translation and x-scaling. However, certain differences in y-scaling and rotation is notable.

Table 2. Transform parameters from frame 15 to frame 180 using registration and the contour tracking algorithm

Method	Translation -X[mm]	Translation -Y[mm]	Scaling -X [100%]	Scaling -Y [100%]	Rotation [rad]	Computing time [ms]
Registration	4.8271	11.8720	1.1139	0.9973	-0.0372	1626640
Contour tracking	4.9212	16.0640	1.0878	0.9845	-0.0195	3.461509

Also, the y-translation in the maximum inhaled and exhaled states proves several millimeters bigger when estimated from the contour tracking algorithm than from registration. Although in the some degrees of freedom there are large differences between the tracking algorithm and image co-registration method, it makes no sense to investigate the difference in each degree of freedom separately, because parameters only make sense as a whole other than considerded as separate numbers. To clarify the effect of the paramter sets on the actual ultrasond images the frame 15 was compared to a later frame 180 which has a large displacement. Respective transform parameters resulting from the contour algorithm and registration are shown in Table 2. To have a visual judgement on the transform quality, the resampled frame 15 with transform parameter sets from each method was subtracted from frame 180, and then the rescaled results are shown in Figure 4. From the figure, we can see that the major displacements around the diaphragm from the original images were corrected with the transform results of both methods. It was also shown that the contour tracking algorithm works better around the vena portae, while the registration yields a superior match at places where the ultrasound probe was contacted to the skin. These differences can mainly be attributed to the fact that the contour tracking

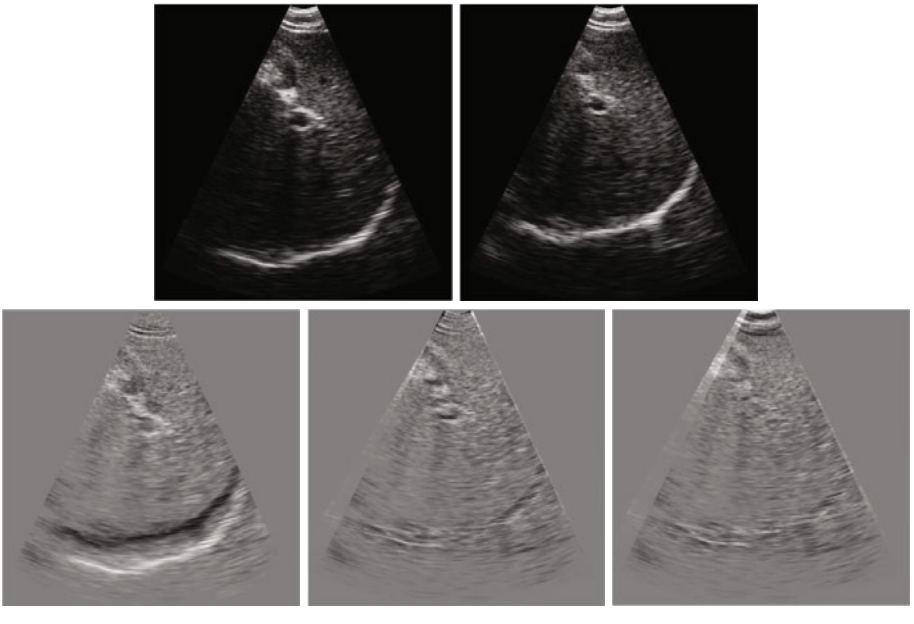


Fig. 4. Upper left: the ultrasound frame 15 in the liver data. Upper right: the ultrasound frame 180 in the liver data. Lower Left: the rescaled subtraction of frame 180 from frame 15. Lower middle: the rescaled subtraction of frame 180 from transformed frame 15 using the result of registration. Lower right: the rescaled subtraction of frame 180 from transformed frame 15 using the result of contour tracking.

algorithm tends to only match the structures a user delineated contour relevant, while the registration takes into account more or less all the image regions.

4 Conclusion

In this work we presented a method to track organ structures on ultrasound images using an algorithm based on conditional density propagation [8]. It was shown that with the method it is feasible to track relevant liver structures real-time on 2D ultrasound images with computing time in the order of milliseconds. The algorithm proves to be robust on low contrast target structures and insensitive against image noise and is therefore well suitable for ultrasound applications. As demonstrated the method provides high reproducibility. Because of the lack of ground truth in vivo tracking, an independent image co-registration registration method was introduced to determine the tracking quality. The comparison shows that the contour tracking algorithm tends to only match the structures a user delineated contour relevant, while the registration takes into account more or less all the image regions, which can be the main reason of the differences between these two methods. As the tracking algorithm leads to superior displacement correction in structures that was previously delineated as part of the contour, a good tracking of the target in real-time can be yielded with delineation of relevant structures on the image and using the suggested algorithm.

This ultrasound tracking method can then be beneficially used in several medical applications eg. the compensation of organ movements in radiation therapy to improve the accuracy of dose delivery to the target, and the real-time reduction of motion artifacts in diagnostic imaging modalities such as CT and MRI. For further clinical utilization, we are now developing the 3D tracking algorithm based on this proposed method.

References

1. Noble, J.A., Boukerroui, D.: Ultrasound image segmentation: A survey. IEEE Trans. Med. Imag. 25(8), 987–1010 (2006)
2. Chalana, V., Linker, D., Haynor, D., Kim, Y.: A multiple active contour model for cardiac boundary detection on echocardiographic sequences. IEEE Trans. Med. Imag. 15(3), 290–298 (1996)
3. Ye, X., Noble, J., Atkinson, D.: 3-d freehand echocardiography for automatic left ventricle reconstruction and analysis based on multiple acoustic windows. IEEE Trans. Med. Imag. 21(9), 1051–1058 (2002)
4. Vitrani, M.A., Ortmaier, T., Morel, G.: Automatic guidance of a surgical instrument with ultrasound based visual servoing. In: IEEE International Conference on Robotics and Automation (ICRA), p. 510 (2005)
5. Hsu, A., Miller, N.R., Evans, P.M., Bamber, J.C., Webb, S.: Feasibility of using ultrasound for real-time tracking during radiotherapy. Med. Phys. 32(6), 1500–1512 (2005)
6. Wein, W., Cheng, J.-Z., Khamene, A.: Ultrasound based Respiratory Motion Compensation in the Abdomen. In: MICCAI 2008 Worshop on Image Guidance and Computer Assistance for Softissue Intervetions, vol. 32(6) (2008)

7. Isard, M., Blake, A.: Contour tracking by stochastic propagation of conditional density. In: Proc. European Conf. on Computer Vision, pp. 343–356 (1996)
8. Isard, M., Blake, A.: Condensation - conditional density propagation for visual tracking. Int. J. Computer Vision 29, 5–28 (1998)
9. Guenther, M., Feinberg, D.A.: Ultrasound-guided MRI: Preliminary results using a motion phantom. Magn. Reson. Med. 52, 27–32 (2004)
10. Ibanez, L., Schroeder, W., Ng, L., Cates, J.: The ITK Software Guide, pp. 215–290. Kitware, Inc. (2003), ISBN 1-930934-10-6

A Malignant Breast Carcinoma Size Assessment Using Multiple Orientation Axial, Lateral, and Shear Elastographies: The Second Stage of a Pilot Study

Ying Chi[1], Michael J. Brady[1], Ruth E. English[2], Junbo Li[1], and J. Alison Noble[1]

[1] Institute of Biomedical Engineering, Department of Engineering Science, Old Road Campus Research Building, University of Oxford, Headington, OX3 7DQ, Oxford, UK
[2] The Churchill Hospital, Oxford Radcliffe Hospital NHS Trust, Old Road, Headington, OX3 7LJ, Oxford, United Kingdom
ying.chi@eng.ox.ac.uk, {jmb,larry,noble}@robots.ox.ac.uk, ruth.english@orh.nhs.uk

Abstract. Elastography is the imaging modality focusing on detection of stiffness variation within inhomogeneous soft tissue. We extended a conventional 2D axial radio frequency (RF) spline correlation based elastography method to multiple orientations, and strengthened its power by equipping it with shear-direction gradient options. This algorithm enhancement explained the elastography from multiple dimensions along lateral and angular orientations, and thus improved its capability in distinguishing tumours from their surrounding desmoplastic fibres. 26 breast cancer cases, from 24 out of 83 recruited patients in our previous study, were re-evaluated, and few Phantom cases were introduced to describe the foundation of the proposed method as benchmarks. Results obtained using the new technique showed strong improvement over the previous method on tumour sensitivity, specificity, fibre recognition and boundary size assessment accuracy.

Keywords: 2D multiple orientation Ultrasound Elasticities, Axial, Lateral, and Shear Strains, malignant breast cancer, size assessment.

1 Introduction

Elastography is an imaging method, first introduced by Ophir et al in 1991 [1], by giving inhomogeneous tissue a quasi-static axial compression and computing the partial derivative of ultrasound-detected tissue element displacements along the axial orientation, indicating tissue stiffness. As cancerous elements are normally stiffer than surrounding normal tissue, this method can be applied to detect cancers. Furthermore, it has been attested to be very promising in distinguishing malignant tumours from the benign [2] and is possible to replace traditional clinical palpation in breast cancer detection in the future. However, because axial elastography acts only along the axial orientation, it is inherently limited in probing 3D targets.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 295–304, 2010.

As the target objects all have a certain 3D geometry, lateral displacement during the quasi-static axial compression (perpendicular to the axial orientation) is normally unavoidable and were first seriously studied and published by Konofagou and Ophir in 1998 [3]. The RF signal amplitude and phase information were interpolated along lateral orientation for displacement and strain computation, and was named lateral elastography. This method compensated for axial strain by considering unconfined volume element lateral slips, which are very common in clinical cases where soft tissue is not normally confined for compressions.

Both axial and lateral strains are designed to demonstrate the central part of a stiff lump (where the displacements are the largest), rather than the boundary conditions of it (which are indicated only by fuzzy gradient changes). In 2007, Thitaikumar, Ophir et al visualized the bonding at an inclusion boundary using axial-shear strain elastography [4]. Patterns from both loosely and tightly bounded inhomogeneous elements were classified, which clinically correspond to malignant and benign tumour boundary conditions respectively [5].

Our objective is to use elastography to determine 2D tumour size assessment, which includes the interests on both cancer volume and boundary visualization. The previous stage of this study focused on the axial strain pattern distribution in 90 breast cancer ultrasound data from 87 patients, which started in 2005 [6, 7]. Axial strain is the most effective of strain types because its orientation of detection is in line with the main displacement. However, in the vicinity of the left and right vertical boundaries of the tumourous tissue, where the transducer is less sensitive to displacement difference among inhomogeneous tissues than to those close to horizontal boundaries, other modalities of strain detection, like lateral & shear-strain elastographies which also consider tissue slipping or tissue property in other orientations, can help. A multiple orientation strain is therefore an obvious enhancement to supply more layers of information to depict the complex volume and boundary situation of cancer.

In this second stage, 26 cases from the original datasets have been restudied by looking into and extracting lateral, axial-shear, and mathematical angular ($M\theta$) strain

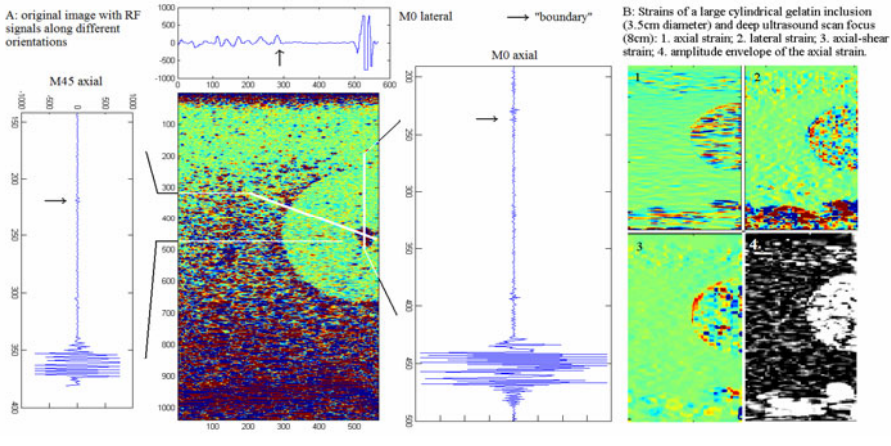


Fig. 1. Examples of ultrasound RF signals, of a 4-fold harder gelatin circular inclusion, along different orientations (*A on the left*), and its *M0* strains (*B on the right*)

patterns (Fig.1A illustrates interpolated RF signals along axial, lateral, and *M45* axial directions of a gelatin cylinder, included inside a softer gelatin environment, here, θ is in degree), which refined our understanding of tissue stiffness distribution by also considering orientation dimensions other than axial elastography. Improvements are shown in nearly all results by comparing them with pathological scans. In many cases ($\geq 85\%$), the desmoplastic fibres were distinguished from the tumourous elements.

Section 2 will introduce the method we used from the aspects of motion estimation, strain extraction, and process control. Section 3 will show results with statistical analysis. Section 4 will compare the proposed method with the other main method of elastography, (i.e. using phase shift technique), briefly digging their pros and cons. Section 5 is a brief update on the ongoing experiments concerning building 3D context for a better 2D elastography understanding, which can be a solid stepping stone between 2D and 3D studies.

2 Method

In a similar way as for the lateral strain, we interpolated the RF signals to splines along $M\theta$. We started this multi-orientation study from θ =30 and θ =45. The θ is called mathematical rotation angle because it can be converted to an actual angle according to specific ultrasound image dimensions (mainly decided by the ultrasound focal depth used), e.g., in Fig.1, the actual axial strain detection orientation is along -19.81° and lateral elastography along 70.19°, when θ is 45.

Based on our first stage study, a spline-based algorithm for continuous time-delay estimation proposed by F Viola [8] was adopted for motion estimation because of its good performance on de-noising jitters and bias using a smooth piecewise cubic spine fitting function to prepare RF data. For strain extraction, since the direct gradient operation normally amplifies image noise, a Least-Square Strain Estimator (LSSE) for Elastography proposed by F Kallel and J Ophir [9] was adopted to reduce noise artifact.

2.1 Displacement Detection

Time Delay Estimation (TDE) is convenient in our study to determine the relative time shift between the pre-compression (reference) signal $s_1(t)$ and post-compression (delayed) signal $s_2(t)$, where t is the time parameter. Both the reference and the delayed signals are frequently seen in a discrete form as

$$s_1(t) = \sum_{i=1}^{M} s_1[i], \quad s_2(t) = \sum_{i=1}^{N} s_2[i], \qquad \text{M} \leq \text{N} \tag{1}$$

In order to find out the relative displacement between $s_1[i]$ and $s_2[i]$, a classic sum square error (SSE) pattern-matching function as equation (2) is convenient to conduct a rigid registration between this image pair.

$$\varepsilon(t) = \sum_{i=1}^{M} (s_2[i] - s_1[i])^2 \tag{2}$$

To obtain good performance on de-noising among time delayed signals, a smooth piecewise cubic interpolation was first applied to $s_2(t)$ (for 2D data, this is conducted by a cubic spline interpolation along axial direction, followed by a same interpolation along the lateral orientation), and this signal after interpolation is represented by $\hat{s}_2(t)$. Since the i^{th} signal of the continuous $\hat{s}_2(t)$ can be written as $f_i(t) = a_i t^3 + b_i t^2 + c_i t + d_i$, the derivative of $\varepsilon(t)$ with respect to t, ie., the Jacobian error shown in equation (3), was computed as an nonlinear optimization scheme to find the centre pixel location of the best match area in $\hat{s}_2(t)$ to the centre point of $s_1[i]$.

$$\begin{aligned}\frac{d\varepsilon(t)}{dt} &= \sum_{i=1}^{M} 2f_i(t)\frac{df_i(t)}{dt} - 2s_1[i]\frac{df_i(t)}{dt} \\ &= t^5\sum_{i=1}^{M} 6a_i^2 + t^4\sum_{i=1}^{M} 10a_ib_i + t^3\sum_{i=1}^{M}(8a_ic_i + 4b_i^2) + t^2\sum_{i=1}^{M}(6a_id_i + 6b_ic_i - 6a_is_1[i]) \\ &+ t\sum_{i=1}^{M}(4b_id_i + 2c_i^2 - 4b_is_1[i]) + \sum_{i=1}^{M}(2c_id_i - 2c_is_1[i])\end{aligned} \quad (3)[8]$$

where $s_1[i]$ has no dependence on t. When equation (3) equals to zero, the SSE is minimized and the 2D distance between $f_i(t)$ and $s_1[i]$ central points are the displacement of $s_2[i]$ along axial and lateral directions.

Since spline correlation for displacement detection on large ultrasound images is normally expected to be of high accuracy but is also computationally intensive, 3-layer Gaussian-window (*Gw*) downsampled-to-fine scaling pyramid were used to convolute $s_1(t)$ and $s_2(t)$ as shown in equation (4). So that, a global-to-local best signal matching procedure was conducted, combining accuracy with efficiency.

$$s_1(t) \approx s_{1Gw}(t) = \sum_{i=1}^{M}\sum_{j=1}^{3} Gw^j \cdot s_1[i], \quad s_2(t) \approx s_{2Gw}(t) = \sum_{i=1}^{N}\sum_{j=1}^{3} Gw^j \cdot s_2[i] \quad (4)$$

We found out that the continuous $f_{ij}(t)$, ie., the cubic spline interpolation on j^{th} Gw downsampled $s_2[i]$, can be equally expressed in a set of divided, independent 2D spline blocks $\sum_{k=1}^{K_j} f_{ijk}(t)$, with approximately 5-pixel margin width along each block side removed in avoiding insignificant marginal difference. K_j is promotional to the downsampled signal size. Then, to implement spline fitting along $M\theta$, $s_{2Gw}[i]$ was broken into a set of pitches $\sum_{k=1}^{K} s_{2Gw}[ik]$, and all these pitches were rotated simultaneously by $M\theta$, before 2D cubic spline interpolation. For convenience and the best signal to noise ratio (SNR), those pitches were centred at the pixels which also calibrate the 3-layer *Gw* pyramid to obtain the same result resolution, as our previous study did, for each strain type. Therefore, the equation (2) was updated as

$$\varepsilon_\theta(t) = \sum_{i=1}^{M}\sum_{j=1}^{3}\sum_{k=1}^{K_j}(\Re_\theta * s_{1Gw^j}[ik] - f_{ijk\theta}(t))^2, \quad (5)$$

where $\Re_\theta$ is the 2D rotation matrix, and $f_{ijk\theta}(t)$ is the interpolated signal upon rotated post-compression signal $s_{2Gw^j}[ik]$.

Both source and target image blocks were rotated in pairs with a sufficient radius to both cover the whole image area in $M\theta$, and to maintain proper overlap areas between blocks for an accurate signal correlation. Thus, rotation windows were superimposed onto the correlation windows, with the minimal rotation radius $r_{\min}$ and correlation window height $h_{\theta,\min}$ and width $w_{\theta,\min}$ given in equation (6).

$$\begin{cases} if\left(h_{M0} \geq w_{M0}\right), \quad h_{M\theta,\min} = 1/\cos(\theta\pi/180)\cdot h_{M0}, \quad w_{M\theta,\min} = \cos(\theta\pi/180)\cdot w_{M0}, \\ else, \quad w_{M\theta,\min} = 1/\cos(\theta\pi/180)\cdot w_{M0}, \quad h_{M\theta,\min} = \cos(\theta\pi/180)\cdot h_{M0}, \\ r_{\min} = \sqrt{(h^2_{M\theta,\min} + w^2_{M\theta,\min})}, \end{cases} \tag{6}$$

where h_{M0} and w_{M0} are predefined window size in our previous study for best detection sensitivity with lowest noise level. Correlation best matching centre pixels in θ were recorded at the same axes location as where they were in our previous study, so that results can be compared without registration.

2.2 Strain Extraction

Tissue strain is the gradient of signal displacements along detection axes within the time interval. Let v and u be the displacements along y and x axes, the strains are

$$\zeta_{x,y} = \partial v/\partial y + \partial u/\partial x + \partial u/\partial y + \partial v/\partial x, \tag{7}[5]$$

where the terms are axial, lateral, axial-shear (AS), and lateral-shear (LS) strains.

For each strain type, LSSE interpolates the M (16 in our study) displacement values, achieved from the displacement detection stage, into one final set of values for strain estimation.

When the applied tissue compression is small (in our study, 3.75mm), the tissue movements are mainly linear. Thus, for each discrete signal $s_2[i]$,

$$\begin{cases} u_i = a_u(z_i \cdot \cos\theta) + b_u \\ v_i = a_v(z_i \cdot \sin\theta) + b_v \end{cases}, \qquad \text{for} \quad i = 1,\ 2,\ ...,M \tag{8}$$

where z_i is the tissue depth, a and b are constants to be estimated for each strain type, and $a + b/z_i$ forms the strain matrix. When b/z_i is small enough to be ignored, a is the maximally approximated strain value.

A quick summarization of this idea is to ascertain the motion state by computing LSSE of the obtained displacements to map a strain pattern which can help predict their future motion tendency when another pressure is applied.

To further increase SNR, the elements gained smaller normalized signal correlation errors (NSCE) in displacement detection were given higher weights in strain extraction. The weight values were computed using the reciprocal of the NSCE as

$$w_i = 2\left\|[v_i, u_i]\right\|/(s_1[i].\text{^}2 + s_2[i].\text{^}2) \tag{9}$$

Thus, using this proposed method, for each axial case in our previous study, 7 more elastography patterns (*M0* lateral and shear strains, $M\theta$ axial, lateral, and their shear strains) were produced for contribution analysis.

2.3 Orientation Selection and Its Process Control

Theoretically, the rotation angle θ can be along any orientation, which should be followed by downsampling scale rate changes to avoid changing the actually implemented rotation angle and to satisfy the resolution need for motion detection along every axis to supply both accuracy and efficiency. The value used in this study for image size of 5×3cm (656×768 in pixel) are:

$$\text{for } M0,\ Gw_y : Gw_x = 2\ ; \quad \text{for } M\theta\ ,\ Gw_y : Gw_x = 1\ . \tag{10}$$

To prevent tissue from slipping outside the 2D scanning plane, tissue decompression was used to collect $s_1(t)$ and $s_2(t)$. Since tissue displacements and slipping orientations were mechanically stimulated by not only the compression force but also the tissue internal structure which were not fully demonstrated in static B-mode image because of its low resolution, slipping during compression process was hard to predict. Decompression is the reverse progress of a compression. During a tissue compression-decompression cycle, clinical doctors first observe the tissue behaviours in compression progress through dynamic B-mode and adjust the compression force till the tissue motions are satisfying, then we recorded the M set of $s_1[i]$ and $s_2[i]$ pairs during the immediately following decompression progress.

Another challenge we also considered was large tumour detection (greater than 15mm in radius). To illustrate this, we used a 3.5cm-diameter gelatin cylinder inclusion, larger than any instruction case in literature. We observed during decompression, that although the inclusion is homogeneous, displacement of elements inside it was not smoothly consistent. As a consequence, the strain pattern inside the inclusion in deep focal depth (e.g., 8cm instead of 5cm from the previous study) exhibits smooth jitters, e.g., the *M0* strains in Fig.1B and Fig.1C. Especially for freehand compression and ultrasound scanning, any tiny hand tremble of the operator can change the element motion speed and cause this problem. However, after observation of around 10 freehand cylindrical phantom results applying various displacement forces on different stiffness inclusions, we discovered the amplitudes of these strain "pulses" can be smoothed by using larger motion detection and/or larger strain extraction windows (e.g., the smooth Fig.2G using strain extraction window of size 12x12 pixels, in comparison to noisy Fig.2H, using 1e-7x1e-7, which is however sharp to boundary information), or by a post-processing quality refinement – amplitude envelope (AE), given in equation (11). AE was implemented for freehand cases if the operator was not steady, where the $\hat{x}^2(t)$ is the Hilbert Transform of $x^2(t)$.

$$A(t) = \sqrt{x^2(t) + \hat{x}^2(t)} = \sqrt{x^2(t) + j[x(t) * (1/(\pi \cdot t))]} \tag{11}$$

One result examples of this analytic refinement on Fig.1B(1) is shown in Fig.1B(4).

3 Result and Evaluation

26 malignant breast cancer and two phantom cases (the circular inclusion in Fig.1 and a plastic dinosaur phantom introduced and shown in Fig.2) have been compared and studied. Phantom data were acquired freehand using a Zonare z.one ultra Convertible Ultrasound System and an 8.5 MHz L10-5 Linear Array 2D transducer. The clinical data were acquired using an Analogic AN2300 digital ultrasound engine which combined a 64-channel digital beamformer front end and a B-K medical 12 MHz 8805 transducer. To avoid freehand system's operator tremble error, an assisted freehand elasticity ultrasound stepper for breast cancer study was designed in our first stage study [6] to contain the transducer, supplying motor controlled automated measureable external compression to reduce movement variability. The ultrasound scanning was also automated and multi-triggered by the motor moving depths. The increased motion control refined image quality and enhanced imaging repeatability. As a radiologist can manually locate it to any lesion area, it also remains sufficiently flexible for breast assessment whilst accommodating these mechanical improvements. For each clinical case, 3.75mm compression was applied, 16 slides were scanned, and the strain extraction window sizes used were 6x3 pixels for $M\theta$ strains, 0.1x3 for *M0* lateral strain, and 3x1 for *M0* axial, AS, and LS strains.

The results of this method were compared to pathological scans, which was supplied by cellular pathologist, following NHS guidance on recognizing cancer cells behind microscopy. 83 patients were recruited, from whom 90 elasticity studies were acquired. Patients requiring a mastectomy, instead of a lumpectomy wide local excision, were excluded in the middle of the study because of the technical problems in the pathological evaluation. This left 31 patients with pathological scans, and 5 of them with pathological scans not showing the ultrasound scan planes (which is a common 2D evaluation difficulty), thus, 26 cases from 24 patients were finally included in the result evaluation. Since very few benign cases were presented, no comparison between malignant and benign was conducted, and all results mentioned here were derived from malignant tumour cases. Result example from a freehand tiny (2%) decompression of a confined 4-fold harder circular inclusion in gelatin is illustrated in Fig.1B. Since our strain computation is based on linear assumptions, smaller movements normally yield better results. The lateral strain image Fig.1C is equally good as the axial strain counterpart in Fig.1B(2) because of its even tinier and stable displacement, or say, lateral slip. The ideal range of tissue movement depends on transducer sensitivity and the use of other equipments to hold transducer steadily to avoid jittering noise. Fig.1B(3) shows the axial-shear strain of Fig.1B phantom, in which the boundary of inclusion has been obviously enhanced. Four clinic result examples are shown in Fig.3. The central tumour in red (T) and surrounding similar-colored fibres (F) are marked in the *M0* axial strain images Fig.3B. The *M0* axial strain excellently detects the upper and lower boundaries of the tumour, while the weighted root mean square (wRMS) sum of *M0* and $M\theta$ axial strains (Fig.3C) supplies more left and right boundary information, and the final strain map in Fig.3D, using weighted sum of all strains from both *M0* and $M\theta$ orientations, separates the tumour from adjacent fibres (weights used are in the range of 0~1.8). Pathological scans (Fig.3E) made after surgery by cellular pathologists were used for evaluation.

Clinical sensitivity and specificity were used as the common evaluation methods for diagnostic tests. The ratio of desmoplastic fibre in the detected tumour area (FIT) was checked to compare methods' capabilities in distinguishing fibres from malignant tissue, where the desmoplastic fibres were defined as the areas wrongly considered as being tumourous by *M0* axial strain. Symmetric boundary Euclidean distance root mean square (SBDRMS) error was employed to evaluate cancer size assessment accuracy. Normalized cross-correlation (NCC) was utilized to relate strain pattern inside the detected tumour to reference texture. Table 1 summarizes the averaged values from these analyses. Sensitivity and signed SBDRMS values echo the consensus that the object size determined using the conventional method is normally greater than the reference size, while our proposed multi-orientation method supplies a more accurate boundary, which can be slightly smaller than the reference, evidenced by specificity, NCC, SBDRMS, and signed SBDRMS. Fig.4 draws the polynomial of FIT ratios from the results using only *M0* axial strain, *M0* strains (including lateral and shear strains), and our proposed $M\theta$ (here, θ =30 to build an actual approximate 45º rotation) strains assisted method. In comparison to the conventional spatial cross-correlation axial elastography, our proposed method demonstrates better performance.

4 Discussion on Comparison with Real-Time Phase-Shift Method

Phase shift is the other mainstream technique of elastography, mainly applied to real-time applications. It is fast due to using only phase spectrum correlation instead of full RF signal spatial cross-correlation. The accuracy of the conventional phase shift method is limited because of few variances among phase spectrums, and aliasing will cause errors in long displacement distance greater than a half wavelength. Both of them mean large motion will easily cause mismatching. Later versions with certain fast correction approaches can be reasonably reliable. For comparison, we applied a real-time method prepared by Shiina et al [10]. It is based on classic Doppler blood flowmetry, enhanced by selecting only nonalias zones for correlation and the estimated distance double confirmed using a smoothed envelope autocorrection. During comparison, we observed that although large displacement will cause blurry boundaries, it is reliable in obtaining global volume representation. However, too strong smoothing degrades its accuracy. Fig.2 demonstrates a comparison example of this methods (Fig.2I) with our proposed method (Fig.2H) on a complex-shape phantom case using a soft-plastic dinosaur which has 30º~45º wings as the gelatin inclusion.

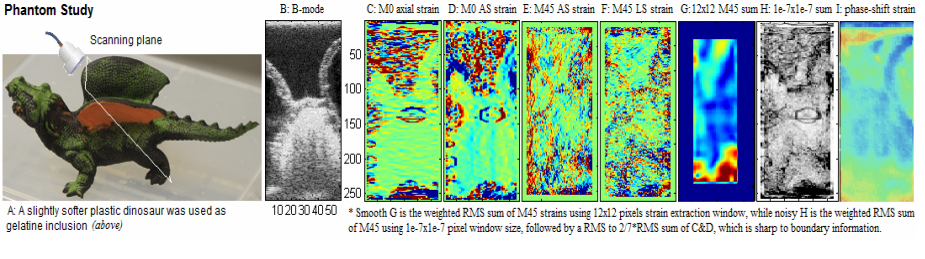


Fig. 2. *M0* and *M45* multi-mode strains (*C~H*), compared to a phase-shift method (*I*)

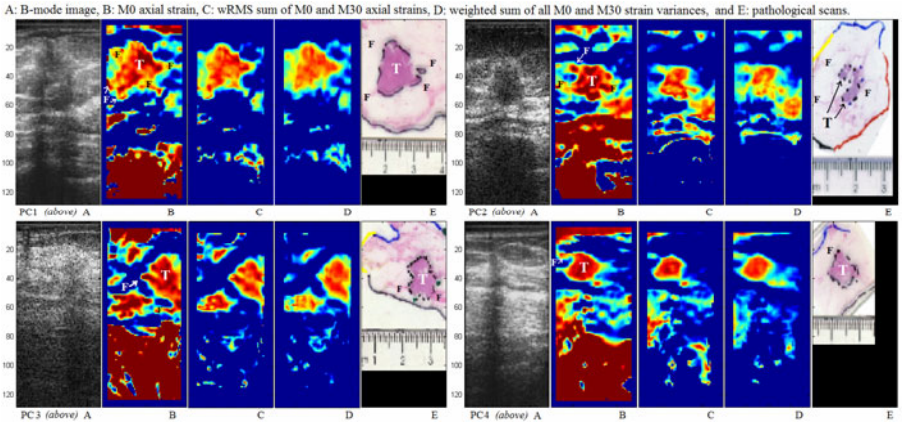


Fig. 3. Four patient case *(PC1~4)* examples illustrate the capability of our proposed method in isolating tumours (areas in red marked T in *B&E*) from desmoplastic fibres (similar-colored parts marked F), evaluated using pathological scans *(E)* from cellular pathologists

Table 1. Statistic summary on clinical and phantom results. As the result of the phase shift method on dinosaur phantom has lower specificity value than *M0* axial strain, its FIT ratio must be larger than that of *M0* axial strain, and thus the computation is waived

Average values	Clinical cases			Dinosaur phantom cases			
	M0 axial strain	wRMS sum of all M0 strains	Weighted sum of all M0 and M30 strain RMSs	M0 axial strain	wRMS sum of all M0 strains	wRMS sum of all M0 and M30 strains	Phase shift method
Sensitivity (%)	76.01%	63.85%	56.74%	84.32%	78.38%	79.06%	80.02%
Specificity (%)	94.01%	96.40%	97.93%	76.02%	82.59%	84.64%	74.40%
FIT (%)	116.82%	71.68%	37.61%	30.20%	20.11%	16.90%	n/a
NCC	0.3862	0.4744	0.5285	0.5503	0.5847	0.5899	0.5358
SBDRMS	6.0236	4.8444	3.1773	2.4821	2.4162	2.3638	3.8516
signed SBDRMS	4.8571	3.0191	0.4710	0.9825	0.1509	0.2425	2.3954

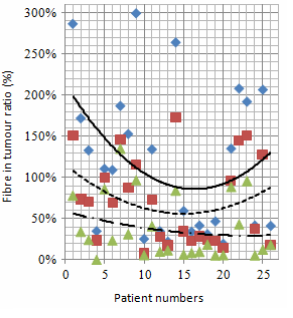
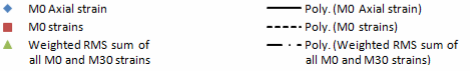


Fig. 4. Polynomial of FIT ratio in clinical study, with the averaged value shown in Table 1



The shape of this inclusion composes of sharp corners and segments with various curvatures, simulating possible malignant tumour shapes. Fig.2C~H illustrates strain types along different orientations. For those $\theta \neq 0$, the strain extraction process can also be rotated like in motion estimation stage for slightly more harmonized result signals. However, the tradeoff will be too focusing on $M\theta$ and losing consideration to *M0* direction factors, rendering the similar problem that *M0* axial strain has.

5 Conclusion and Future Work

This paper presented the second stage of a clinical pilot study, which analyzes malignant cancer ultrasound scans from extended rotational dimensions using multi-orientation and

multi-model elastography. Distinguishing the desmoplastic fibre from the tumour is no longer impossible.

To continue this study, circumference strain pattern and resolution analysis are currently being investigated. Stitching 3D context is also crucial to a better 2D-elastography pattern understanding, especially when the target object is larger than the 2D transducer capture range. Many traditional feature-extraction based registration methods can fail from mistakes in choosing proper matching or aligning features amongst all image properties. Full intensity based 3D NCC volume registration is also under testing, from different angles and detecting locations, to assist elastographic studies. The registration methods currently under scrutiny include pairwise registration, global registration, narrow-volume registration, all the way towards tissue stitching or mosaicing. This registration study can be a safe stepping stone where the research progress can be pushed from 2D to 3D and 4D.

Acknowledgments. This work is supported by the Oxford Biomedical Research Centre (Imaging Theme).

References

1. Ophir, J., Céspedes, I., Ponnekanti, H., Yazdi, Y., Li, X.: Elastography: A Method for Imaging the Elasticity of Biological Tissues. Ultrasonic Imaging 13(2), 111–134 (1991)
2. Itoh, A., Ueno, E., Tohno, E., Kamma, H., Takahashi, H., Shiina, T., et al.: Breast Disease: Clinical Application of US Elastography for Diagnosis. Radiology 239(2), 341–350 (2006)
3. Konofagou, E., Ophir, J.: A New Elastographic Method for Estimation and Imaging of Lateral Displacements, Lateral Strains, Corrected Axial Strains and Poisson's Ratios in Tissues. Ultrasound in Med. & Biol. 24(8), 1183–1199 (1998)
4. Thitaikumar, A., Krouskop, T.A., Garra, B.S., Ophir, J.: Visualization of Bonding at an Inclusion Boundary using Axial-Shear Strain Elastography: a Feasibility Study. Phys. Med. Biol. 52, 2615–2633 (2007)
5. Thitaikumar, A., Mobbs, L.M., Ophir, J., et al.: Breast Tumour Classification Using Axial Shear Strain Elastography: a Feasibility Study. Phys. Med. Biol. 53, 4809–4823 (2008)
6. Kadour, M.J., Noble, J.A.: Assisted-freehand ultrasound elasticity imaging. IEEE Trans. on UFFC 56(1), 36–43 (2008)
7. Li, J., Cui, Y., Noble, J.A.: Ultrasound-based estimation of breast tissue biomechanical properties. J. Strain Analysis 44, 363–374 (2009)
8. Viola, F., Walker, W.F.: A Spline-Based Algorithm for Continuous Time-Delay Estimation Using Sampled Data. IEEE Trans. Ultrason. Ferroelectr. Freq. Control. 52(1), 80–93 (2005)
9. Kallel, F., Ophir, J.: A Least-Squares Strain Estimator for Elastography. Ultrasound Imaging 19, 195–208 (1997)
10. Shiina, T., Nitta, N., Ueno, E., Bamber, J.C.: Real Time Tissue Elasticity Imaging Using the Combined Autocorrelation Method. J. Med. Ultrasonics. 26(2), 57–66 (1999)

Level Set Diffusion for MRE Image Enhancement

Bing Nan Li[1,6], Chee Kong Chui[2], Sim Heng Ong[3,4], Stephen Chang[5], and Etsuko Kobayashi[6]

[1] NUS Graduate School for Integrative Science and Engineering,
[2] Department of Mechanical Engineering,
[3] Department of Electrical and Computer Engineering,
[4] Division of Bioengineering,
National University of Singapore, Singapore
[5] Department of Surgery, National University Hospital, Singapore
[6] Department of Precision Engineering, University of Tokyo, Tokyo, Japan
bingoon@ieee.org, {mpecck,eleongsh}@nus.edu.sg,
etsuko@bmpe.t.u-tokyo.ac.jp

Abstract. Magnetic resonance elastography (MRE) is an emerging technique for noninvasive imaging of tissue elasticity. Proprietary algorithms are used to reconstruct tissue elasticity from the images of wave propagation within soft tissue. Elasticity reconstruction suffers from interfering noise and outliers. The interference causes biased elasticity and undesired artifacts in the reconstructed elasticity map. Anisotropic geometric diffusion is able to suppress image noise while enhance inherent features. Therefore we integrate anisotropic diffusion with level set methods for numerical enhancement of MRE wave images. Performance evaluation of the proposed level set diffusion (LSD) approach was conducted on both synthetic and real MRE datasets. Experimental results confirm the effectiveness of LSD for MRE image enhancement and direct inversion.

Keywords: Anisotropic diffusion, image enhancement, level set methods, magnetic resonance elastography (MRE), noise suppression.

1 Introduction

Magnetic resonance elastography (MRE) is an emerging technology for noninvasive imaging of tissue elasticity [1,2]. It has been evaluated for liver fibrosis, brain degeneration, muscular activity, and others [1-3]. MRE does not characterize tissue elasticity directly. Mechanical properties have to be estimated from the phase images of wave propagation. Several reconstruction algorithms, including local frequency estimation [4] and algebraic inversion of Helmholtz equation [5,6], have been proposed to analyze MRE wave images. Their performance against synthetic Gaussian white noise was examined in [2,6]. The authors claim that most reconstruction algorithms are robust to moderate noise, but incur biased elasticity and undesired artifacts. It is noteworthy that the study in [2,6] is relevant to Gaussian white noise only.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 305–313, 2010.

Nevertheless, we find that MRE noise cannot be simply characterized with a Gaussian distribution. Other investigators also pointed out that the noise in magnetic resonance imaging (MRI) is often subjected to speckle or Rician distributions [7,8]. A simple median filter has been evaluated for noise suppression in [2,6]. However, median filters are optimal for isolated outliers only [8,9].

Both Gaussian and median filters are based on statistical analysis of image pixels. In contrast, anisotropic diffusion carries out noise suppression according to inherent geometric features. There are many studies motivated by the seminal works of Perona and Malik in [10]. Malladi and Sethian put geometric curvature flows in a level set framework for unified noise suppression and image enhancement [11]. Weickert introduced the concept of structural tensor into anisotropic diffusion in order to enhance flow-type features [12]. Gilboa extended this framework to complex diffusion [13]. A distinct advantage of anisotropic geometric diffusion lies in feature preservation during noise suppression. Its potentials in medical image processing were demonstrated in several papers [11,12,14,15].

It is not easy to obtain ideal wave patterns with sufficiently good signal-to-noise ratio (SNR) by means of MRE [6,16]. There are always uncontrollable noise and outliers from mechanical actuators and/or MRI that cannot be suppressed by phase difference. Furthermore, we find that both Gaussian and median filters, which are widely employed in MRE analysis [2,4,6,20], are not efficient to suppress those refractory noise and outliers. In this paper we consider nonlinear geometric diffusion and integrate it into a level set framework for MRE image enhancement. Other than classical anisotropic diffusion, we also investigate the complex model of level set diffusion (LSD) that is particularly suitable for MRE wave field. Its contributions to elasticity reconstruction will be verified qualitatively and quantitatively on synthetic and real MRE datasets.

2 Methods

The propagation of elastic waves in soft tissue is complicated. Multi-variant, multi-dimensional dynamical equations have been proposed to model this phenomenon [17]. Nevertheless, in order to simplify the inversion process, it is reasonable to assume that soft tissues are isotropic, incompressible and possess local homogeneity. The complicated equations then could be decomposed into independent Helmholtz equations governing the harmonic elastic planar waves in different orientations:

$$\nabla \cdot ((\mu + i\omega\zeta)\nabla u_i) = -\rho\omega^2 u_i, \qquad i = 1,2,3. \tag{1}$$

The orthogonal u_1, u_2 and u_3 characterize the displacement or wave field; μ (shear modulus) and ζ (shear viscosity) describe the mechanical properties of soft tissue against shear planar wave; ρ is tissue density; and ω is the steady exciting frequency. With Eq. (1), shear modulus of soft tissue can be estimated by a 2D planar strain wave. In addition to a reduction in the time of MRE imaging, the reconstruction of tissue elasticity is simplified [2,6,16].

The 2D Helmholtz equation in Eq. (1) may be directly approximated by algebraic inversion [2,5,6]:

$$\mu + i\omega\zeta = -\rho\omega^2 \frac{u}{\nabla^2 u}, \tag{2}$$

where u is the planar shear wave in a specific direction, and ∇^2 denotes the geometric Laplacian. MRE imaging attempts to snapshoot the temporal propagation of shear wave with different phase offsets. It is possible to derive the complex wave field in the specific harmonic frequency from the temporal phase images [2,18]. The solution of Eq. (2) is thus a complex distribution. Shear modulus μ and viscosity ζ can be recovered independently from the real and imaginary components. This paper focuses on the shear modulus only.

For discrete wave field, the spatial Laplacian ∇^2 can be approximated conveniently by central difference [19]. However, solving Eq. (2) is often confounded by noisy wave fields [5,6,16]. Matched filters [18] and Savitsky-Golay filters [5] have been proposed as a remedy in this regard. Both of them, nonetheless, are effective for moderate Gaussian white noise only.

2.1 Level Set Diffusion

The conventional low-pass smoothing is not efficient for non-white noise. More advanced methods, such as multispectral analysis and anisotropic diffusion, are necessary in the cases of Rician and/or speckle noise [9,15]. We are particularly interested in nonlinear geometric diffusion because it directly relates to the divergence or Laplacian operator in MRE inversion.

A wave field u_0 may be integrated into a level set framework for geometric diffusion:

$$u_t = c \cdot \vartheta u, \qquad u_{t=0} = u_0, \tag{3}$$

where t is the temporal index of evolution, c is a small coefficient for stable evolution, and ϑ denotes the specific geometric operator, such as divergence [10] or mean curvature [11], on u. In essence, Eq. (3) considers the wave field u_0 as a collection of isocontours with different amplitudes, and evolves them according to their inherent geometries. Such implicit representation facilitates the numerical approximation of involved geometries [19]. For MRE image enhancement, the Laplacian ∇^2 for geometric diffusion is considered.

Although isolated noise and variational boundaries evolve faster, the isotropic diffusion in Eq. (3) removes noise, outliers as well as important patterns away. Eventually it will maintain at a constant amplitude. Therefore it is necessary to control the above geometric diffusion. Perona and Malik proposed a decreasing function of image gradients in order to suppress image noise but preserve strong boundaries:

$$f = (1 + (|\nabla u| / \tau)^2)^{-1}, \tag{4}$$

where τ is a gradient threshold. The geometric diffusion that follows is described by

$$u_t = c\nabla \cdot (f(|\nabla u|)\nabla u), \qquad u_{t=0} = u_0. \tag{5}$$

Such a diffusion model is able to suppress random noise quickly but preserve strong variations for a long time. It is known as the anisotropic diffusion.

Intensity gradients are not robust against noise and artifacts [9,13-15]. A common remedy is to convolve the original image with a low-pass Gaussian kernel [14,15]. By solving a time-dependent Schrodinger equation, Gilboa found that the imaginary field, complementary to the image of interest, is a more reliable approach to control geometric diffusion [13]. His diffusion model relies on the controlling function:

$$f(\mathrm{Im}(u)) = e^{i\theta} \Big/ (1 + (\frac{\mathrm{Im}(u)}{\tau\theta})^2), \tag{6}$$

where 'Im' denotes the imaginary part of the complex field, θ is a phase angle that is necessary to normalize the imaginary field, and τ is a threshold similar to the one in Eq. (4). The complex diffusion model is particularly suitable for MRE image enhancement in that the extracted displacement has been a complex wave field.

Direct inversion of tissue elasticity suffers from noise and outliers in wave fields. It is therefore attractive to investigate the level set-based geometric diffusion for MRE image enhancement. Both complex anisotropic diffusion (CAD) and the classical paradigm by Perona and Malik (PMD) are implemented and evaluated against conventional Gaussian and median filters in this paper.

3 Experiments

We developed a numerical simulator for MRE in Matlab® (MathWorks, USA) using its Partial Differential Equation (PDE) toolbox. A two-dimensional 20cm×20cm phantom (Fig. 1(a)) was generated with the simulator. The boundary condition was configured for full wave reflection. There were two embedded objects with elasticities (green: 2kPa; orange: 4kPa) different from that (1kPa) of the background phantom. The vibrator, exerting a 200Hz planar shear wave, was positioned at the bottom. With a spatial resolution of 256×256, the simulated wave field u was illustrated in Fig. 1(b). The shear modulus of direct inversion by Eq. (2) is shown in Figs. 1(c) & (d).

The experiment was also conducted on a publicly available MRE dataset of agarose gel phantom [20]. The background phantom was 1.5% agarose gel, and the embedded objects were four cylinders, comprised of 10% bovince gel, with different diameters (5~25mm). The excitation frequency was 100Hz. Over a complete wave period, there were 8 snapshots of 256×256 images with different phase offsets. Fig. 2 shows the magnitude image (a), the real part of the extracted wave field (b), as well as the reconstructed shear modulus (c) & (d) using direct inversion.

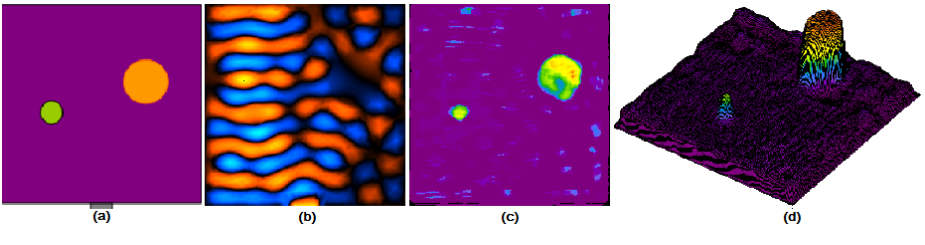


Fig. 1. MRE simulation. (a) Simulation setup; (b) Real part of simulated wave field; (c) Reconstructed shear modulus; (d) Surface rendering of (c).

The impact of Gaussian noise has been examined carefully in [2]. However, as pointed out in [7], it is a common misconception to assume that noise is a zero-mean, constant-variance Gaussian distribution in medical imaging. In fact, the noise is often speckle and dependent on its residence [15]. We synthesized the speckle and Rician noise respectively in order to examine LSD for refractory noise suppression. Given a complex wave image u, the degenerated version with speckle noise was scaled as:

$$u_{\mathrm{sp}} = (|u| + n_s\sqrt{|u|})e^{j(\angle u)}, \tag{7}$$

where j is the square root of -1, $|\cdot|$ denotes the magnitude, $\angle$ refers to the phase, and n_s is independent Gaussian white noise. The wave image contaminated by Rician noise is given by [8]:

$$u_{\mathrm{rc}} = (\sqrt{(|u|/\sqrt{2} + n_s)^2 + (|u|/\sqrt{2} + n_t)^2})e^{j(\angle u)}, \tag{8}$$

where n_s and n_t are independent Gaussian white noise.

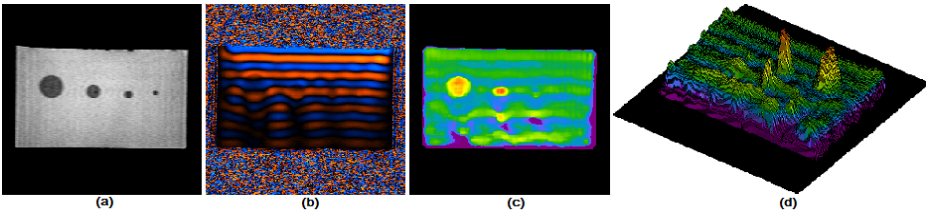


Fig. 2. Real MRE dataset of an agarose gel phantom. (a) Magnitude image; (b) Real part of wave field; (c) Reconstructed shear modulus; (d) Surface rendering of (c).

4 Results and Discussion

The authors in [2], by evaluating synthetic Gaussian noise, drew the conclusion that direct inversion is sensitive to noise. The non-white speckle and Rician noise is even more refractory than Gaussian white noise. Fig. 3 illustrates the adverse impacts of noise on direct inversion. The first row is on the simulated elastic wave, while the second row is on the real MRE dataset. The columns (a) and (c) depict the wave fields with speckle and Rician noise; the columns (b) and (d) show the reconstructed elasticity maps respectively. Although speckle and Rician noise has not degraded the wave fields too much, the shear moduli by direct inversion turn out to be too noisy to be understood. It is thus necessary to enhance MRE images before direct inversion.

We implemented the classical Gaussian and median filters for MRE image enhancement. Both filters adopted a 3×3 scoping window, and the Gaussian filter had an additional smoothing coefficient of 2. Fig. 4 illustrates the reconstructed shear moduli after noise suppression. The Gaussian filter led to the results in the first row, while the median filter led to those in the second row. Intuitively, both Gaussian and median filters are not sufficiently robust to suppress non-white speckle and Rician noise.

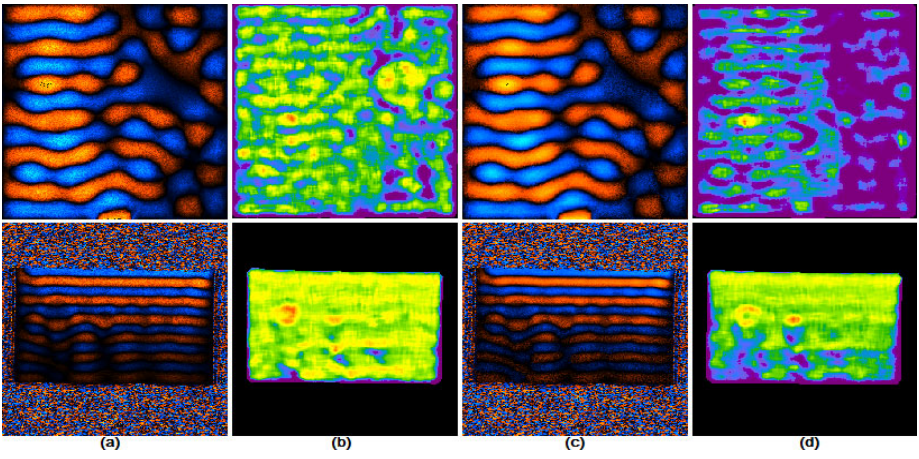


Fig. 3. The impacts of noise on MRE inversion. (a) Wave fields contaminated by speckle noise; (b) Shear modulus by direct inversion on (a); (c) Wave fields contaminated by Rician noise; (d) Shear modulus by direct inversion on (c).

LSD is able to suppress noise while preserving important features. Nevertheless, its performance is subject to appropriate controlling parameters [9,14,15]. For instance, PMD is controlled by the gradient threshold τ and the iteration of diffusion T, and CAD has an additional phase angle θ. The authors in [13,15] have confirmed that θ is necessary for imaginary field normalization, but not critical for diffusion as long as it is kept small. Therefore we fixed θ as an empirical value of $\pi/30$ ($\approx$0.105) in all experiments. The effectiveness of LSD was measured on the wave field itself as well as the estimated shear moduli. The signal-to-mean square error (SMSE) ratio [21] was chosen as the quantitative index for wave fields:

$$\frac{\mathrm{S}}{\mathrm{MSE}} = \log_2\left[\frac{\sum \omega^2}{\sum (\varpi - \omega)^2}\right], \tag{9}$$

where ω denotes the original clean image, and ϖ is the noisy image after enhancement. The larger the value of SMSE is, the better the image enhancement is. We found that, in most cases, a conservative configuration (i.e., τ=0.1 and T=10) has been able to enhance MRE images well.

Fig. 5 shows the results of LSD for MRE image enhancement and direct inversion. The results in the first row were due to PMD, while those in the second row were enhanced by CAD. Although the elastograms in Fig. 3 are nearly illegible, the resultant images after LSD enhancement are clean and visually close to the genuine elastograms. By comparing the results in Figs. 4 and 5, LSD obviously outperforms Gaussian and median filters for MRE image enhancement and direct inversion. Furthermore, CAD performs better than PMD for MRE image enhancement and direct inversion, which is obvious in Figs. 5(a) and 5(b).

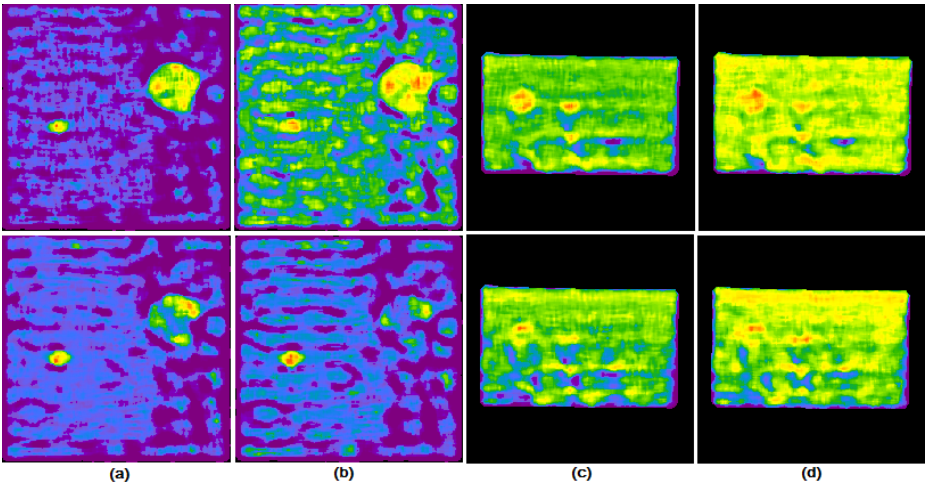


Fig. 4. MRE image enhancement by Gaussian and median filters. The first row is on Gaussian smoothing, while the second row is on median filtering. (a) & (c) Direct inversion after suppressing speckle noise; (b) & (d) Direct inversion after suppressing Rician noise.

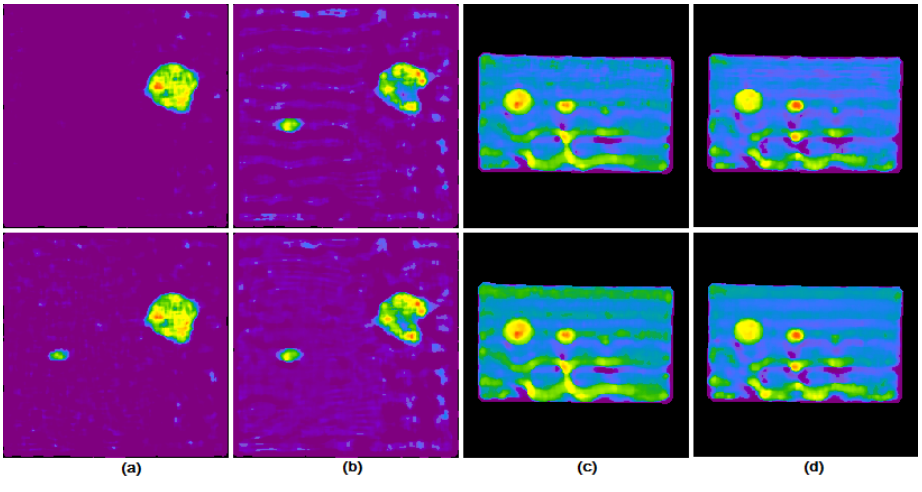


Fig. 5. LSD for MRE image enhancement. The first row is on PMD while the second row is on CAD. (a) & (c) Direct inversion after suppressing speckle noise; (b) & (d) Direct inversion after suppressing Rician noise.

Both intensity consistency and structural similarity are important in medical image processing [21]. Eq. (9) is an index oriented to intensity consistency. We propose the following index for structural similarity (SIM):

$$\mathrm{SIM} = 1 - \left[\frac{\sum(\varpi_b \oplus \omega_b)}{\sum(\varpi_b \vee \omega_b)} \right], \tag{10}$$

where ⊕ denotes logical exclusive-OR, □ means logical OR, and the subscript '$_b$' denotes image thresholding. In other words, merely the regions with distinct shear modulus are preserved in the binary images. SIM approaches to the value 1 if two binary patterns are sufficiently close to each other.

As shown in Table 1, all filters are helpful for MRE image enhancement and noise suppression. Both PMD and CAD are robust to suppress refractory noise and enhance direct inversion. Nevertheless, the effectiveness of Gaussian and median filters varies from case to case. The results in Fig. 5 are obviously more legible than those of Fig. 4, whereas their quantitative indices are close. When there is a need for compromise between intensity consistency and structural similarity, the latter is often preferred in medical image analysis [21].

Table 1. Intensity consistency and structural similarity of the reconstructed elastograms

MRE	Noise	SMSE					SIM				
		Noisy	Gauss	Med	PMD	CAD	Noisy	Gauss	Med	PMD	CAD
Simulation	Speckle	1.25	2.73	2.02	2.04	2.13	0.37	0.81	0.42	0.76	0.82
	Rician	1.32	2.28	1.87	2.57	2.39	0.39	0.55	0.42	0.77	0.78
Real	Speckle	2.73	4.35	3.30	3.58	3.67	0.37	0.62	0.42	0.61	0.59
	Rician	2.06	3.27	2.74	3.54	3.30	0.37	0.46	0.40	0.55	0.55

5 Conclusion

Direct algebraic inversion of Helmholtz equations is an attractive method for elasticity reconstruction from MRE wave images. It is efficient for numerical approximation and robust to moderate noise. However, we find that the real MRE noise is more refractory than Gaussian white noise. The noise is often speckle and close to Rician distribution. It is then difficult to employ direct inversion to reconstruct shear modulus. Noise suppression is necessary for MRE image analysis.

Conventional Gaussian and median filters are not efficient to suppress the refractory speckle and Rician noise. We proposed to use anisotropic geometric diffusion integrated into a level set framework for MRE image enhancement. In the level set framework, we conducted the classic anisotropic diffusion as well as the state-of-the-art paradigm of complex diffusion. Both qualitative and quantitative evaluations confirmed them to be effective for noise suppression. In summary, LSD is a promising approach to suppress MRE noise and enhance direct inversion. Currently we are carrying out MRE imaging on agarose phantoms and animal organs as well. LSD will be further examined and improved for those MRE images.

References

1. Greenleaf, J.F., Fatemi, M., Insana, M.: Selected methods for imaging elastic properties of biological tissues. Annual Review on Biomedical Engineering 5, 57–78 (2003)
2. Manduca, A., Oliphant, T.E., Dresner, M.A., et al.: Magnetic resonance elastography: Non-invasive mapping of tissue elasticity. Medical Image Analysis 5, 237–254 (2001)
3. Dresner, M.A., Rose, G.H., Rossman, P.G., et al.: Magnetic resonance elastography of skeletal muscle. Journal of Magnetic Resonance Imaging 13, 269–276 (2001)

4. Manduca, A., Muthupillai, R., Rossman, P.J., et al.: Image processing for magnetic resonance elastography. In: Proc. SPIE, vol. 2710, pp. 616–623 (1996)
5. Oliphant, T.E., Manduca, A., Ehman, R.L., Greenleaf, J.F.: Complex-valued stiffness reconstruction for magnetic resonance elastography by algebraic inversion of the differential equation. Magnetic Resonance in Medicine 45, 299–310 (2001)
6. Bishop, J., Samani, A., Sciarretta, J., Plewes, D.B.: Two-dimensional MR elastography with linear inversion reconstruction and noise analysis. Phys. Med. Biol. 45(8), 2081–2091 (2000)
7. Gravel, P., Beaudoin, G., De Guise, J.A.: A method for modeling noise in medical images. IEEE Trans. Med. Imag. 23(10), 1221–1232 (2004)
8. Manjon, J.V., Carbonell-Caballero, J., Lull, J.J., et al.: MRI denoising using Non-Local Means. Medical Image Analysis 12(4), 512–523 (2008)
9. Buades, A., Coll, B., Morel, J.M.: A review of image denoising algorithms, with a new one. Multiscale Model. Simul. 4(2), 490–530 (2005)
10. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Trans. PAMI 12(7), 629–639 (1990)
11. Malladi, R., Sethian, J.A.: Image processing via level set curvature flow. Proc. Natl. Acad. Sci. 92, 7046–7050 (1995)
12. Weickert, J.: Coherence-enhancing diffusion filtering. International Journal of Computer Vision 31, 111–127 (1999)
13. Gilboa, G., Sochen, N., Zeevi, Y.Y.: Image enhancement and denoising by complex diffusion processes. IEEE Trans. PAMI 26(8), 1020–1036 (2004)
14. Gerig, G., Kubler, O., Kikinis, R., Jolesz, F.A.: Nonlinear anisotropic filtering of MRI data. IEEE Trans. Med. Imag. 11(2), 221–231 (1992)
15. Salinas, H.M., Fernandez, D.C.: Comparison of PDE-based nonlinear diffusion approaches for image enhancement and denoising in optical coherence tomography. IEEE Trans. Med. Imag. 26(6), 761–771 (2007)
16. Kwon, O.I., Park, C., Nam, H.S., et al.: Shear modulus decomposition algorithm in magnetic resonance elastography. IEEE Trans. Med. Imag. 28(10), 1526–1533 (2009)
17. Graff, K.F.: Wave Motion in Elastic Solids. Dover Publications, New York (1991)
18. Manduca, A., Lake, D.S., Kruse, S.A., Ehman, R.L.: Spatio-temporal directional filtering for improved inversion of MR elastography images. Medical Image Analysis 7, 465–473 (2003)
19. Osher, S., Fedkiw, R.: Level Set Methods and Dynamic Implicit Surfaces. Springer, Heidelberg (2003)
20. Grimm, R.C., Lake, D.S., Manduca, A., Ehman, R.L.: MRE/Wave. Mayo Clinics, Rochester, MN, USA, `http://mayoresearch.mayo.edu/ehman_lab/`
21. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality evaluation: From error visibility to structural similarity. IEEE Trans. Image Processing 13(4), 1–14 (2004)

Content-Based Surgical Workflow Representation Using Probabilistic Motion Modeling

Stamatia Giannarou[1] and Guang-Zhong Yang[1,2]

[1] Institute of Biomedical Engineering
[2] Department of Computing
Imperial College London, SW7 2BZ, UK

Abstract. Succinct content-based representation of minimally invasive surgery (MIS) video is important for efficient surgical workflow analysis and modeling of instrument-tissue interaction. Current approaches to video representation are not well suited to MIS as they do not fully capture the underlying tissue deformation nor provide reliable feature tracking. The aim of this paper is to propose a novel framework for content-based surgical scene representation, which simultaneously identifies key surgical episodes and encodes motion of tracked salient features. The proposed method does not require pre-segmentation of the scene and can be easily combined with 3D scene reconstruction techniques to provide further scene representation without the need of going back to the raw data.

Keywords: surgical workflow analysis, motion modeling.

1 Introduction

Assessment of surgical workflow for Minimally Invasive Surgery (MIS) is valuable for evaluating surgical skills and designing context-sensitive surgical assistance systems. In this regard, direct use of MIS video has many practical advantages as it does not complicate the existing operating room settings. However, the challenges associated with this approach are also evident. MIS video data are associated with high temporal redundancy making off-line analysis of surgical workflow difficult. Being extremely voluminous, they require significant time for visualization. Also, information of interest can be accessed by sequential video scanning and data manipulation and editing is achieved by frame-by-frame video processing. Extensive research has already been conducted for surgical scene segmentation, instrument tracking, tissue deformation recovery and modeling [1].

In practice, the large volume of video data involved raises a significant information management challenge and a succinct content-based representation is required to facilitate efficient indexing, browsing, retrieval and comparison of relevant information for faster and more comprehensive analysis and understanding of the surgical workflow with minimal information loss. This requires a

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 314–323, 2010.

high-level representation of visual information which reflects not only the scene structure but also the underlying semantic and context of the *in vivo* environment. Common approaches to video representation follow the framework of the MPEG-7 standardization [2]. The identification of representative frames, the so called "keyframes", has been used extensively to convey the content of videos [3]. Although useful for broadcasting and general use, such a representation is of limited use for MIS workflow analysis as it does not adequately represent the underlying information and instrument-tissue interaction.

One prerequisite of content-based representation of surgical workflow is the identification of surgical episodes that portray different events within the sequence. Existing work for surgical workflow segmentation has focused on the analysis of surgical actions. Probabilistic frameworks have been used to classify surgical actions based on detecting instrument-tissue interaction by tracking surgical tools and incorporating multiple visual cues [4]. Special MIS tools have also been designed, equipped with force and torque sensors to facilitate the classification of surgical actions by considering instrument kinematics [5]. Dynamic Time Warping (DTW) and Hidden Markov Model (HMM) based approaches have been proposed to classify the overall surgical procedures [6]. Additional information such as eye-gaze and hand/limb movement has also been included to improve the overall sensitivity and specificity of the analysis framework [7].

The purpose of this paper is to propose a new content-based surgical workflow representation scheme that is suitable for both surgical episode identification and instrument-tissue motion modeling. The aim is to transform the MIS data from an implicit and redundant frame-based representation to an explicit, abstracted, high-level surgical episode description. The proposed approach does not require pre-segmentation and the motion characteristics of salient features are used to identify tissue deformation in response to instrument interaction. Surgical episodes can be naturally derived from this framework and further representation of the scene with 3D reconstruction using motion or simultaneous localization and mapping (SLAM) can be readily built on top of this representation without the need of going back to the raw data.

2 Methods

2.1 Episode Segmentation

In this work, surgical episodes describe distinct surgical actions and consecutive episodes represent surgical steps. The visual content information of MIS data is expressed by the action events of salient features. Episode borders occur when feature tracking fails as this is due to the contrastingly different visual appearance of the changing surgical environment.

The affine-invariant anisotropic region detector [8] is employed for reliable and persistent feature tracking where salient features are identified as points that have strong gradients and are anisotropic along several directions. Scale adaptation is based on the strength of the anisotropic pattern whereas affine adaptation relies on the pattern's intrinsic Fourier properties. An Extended Kalman

Filter (EKF) parameterization scheme is used to adaptively adjust the optimal templates of the detected affine-invariant anisotropic regions, enabling accurate identification and matching of a set of tracked features over a series of frames. The information provided to the EKF is the location of a salient point in each frame and the parameters of the ellipse that represent its region. The state of the EKF consists of the coordinates of the ellipse center (x, y), the velocities along the horizontal and vertical axes (u, v), the coordinates of the tip of the major axis (r^x, r^y), the angle between the horizontal and the major axis of the ellipse θ, the angular velocity ω, and the ratio between the major and the minor axes k of the ellipse. The state of a salient point at time t is defined as:

$$s_t = f(s_{t-1}, w_t) = \begin{bmatrix} x_t \\ y_t \\ u_t \\ v_t \\ \theta_t \\ \omega_t \\ r_t^x \\ r_t^y \\ k_t \end{bmatrix} = \begin{bmatrix} x_{t-1} + (u_{t-1} + w_{t-1}^u) \\ y_{t-1} + (v_{t-1} + w_{t-1}^v) \\ u_{t-1} + w_{t-1}^u \\ v_{t-1} + w_{t-1}^v \\ \theta_{t-1} + \omega_{t-1} + w_{t-1}^\omega \\ \omega_{t-1} + w_{t-1}^\omega \\ r_t^x \\ r_t^y \\ k_{t-1} \end{bmatrix} \tag{1}$$

where, $w_t = [w_t^u, w_t^v, w_t^\omega]$ is zero mean additive Gaussian noise and r_t^x and r_t^y are the results of the homography:

$$\begin{bmatrix} r_t^x \\ r_t^y \end{bmatrix} = \begin{bmatrix} \cos(\omega_{t-1} + w_{t-1}^\omega) & -\sin(\omega_{t-1} + w_{t-1}^\omega) \\ \sin(\omega_{t-1} + w_{t-1}^\omega) & \cos(\omega_{t-1} + w_{t-1}^\omega) \end{bmatrix} \cdot \begin{bmatrix} r_{t-1}^x - x_{t-1} \\ r_{t-1}^y - y_{t-1} \end{bmatrix} + \begin{bmatrix} u_{t-1} + w_{t-1}^u + x_{t-1} \\ v_{t-1} + w_{t-1}^v + y_{t-1} \end{bmatrix} \tag{2}$$

The aim of tracking is to establish frame correspondence between region $\hat{s}_t^-$ predicted by the EKF and the detected regions in the search window at the examined frame. In this work, we use the relative amount of overlap O in the image area covered by the compared regions and the dissimilarity C in the anisotropic measure of the compared features as an indication of region correspondence, defined as:

$$O_{A,B} = 1 - \frac{A \cap B}{A \cup B} \quad , \quad C_{A,B} = \frac{|c_A - c_B|}{\max_{n \in S} |c_A - c_n|} \tag{3}$$

where $A \cap B$ and $A \cup B$ are the intersection and the union, respectively of regions A, B, c_A represents the anisotropic measure of feature A and S denotes the search area.

The EKF corrected state $\hat{s}_t^+$ is verified using spatial context and regional similarity. The context of a feature is represented by a set of auxiliary features that exhibit strong motion correlation to the feature and is used to estimate an approximate location $\{\tilde{x}_t, \tilde{y}_t\}$ of the feature. The region similarity is estimated based on the Bhattacharyya distance between their RGB histograms. The spatial context information is also used to boost the prediction of the EKF and recover

potential tracking failure. When $\hat{s}_t^-$ is not able to be matched to any of the detected features or $\hat{s}_t^+$ is not a valid correspondence, the location $\{\tilde{x}_t, \tilde{y}_t\}$ is used to generate a "hypothetical" predicted state, defined as:

$$\begin{aligned}\hat{s}_t^h = [&\tilde{x}_t, \tilde{y}_t, \tilde{x}_t - \hat{x}_{t-1}^+, \tilde{y}_t - \hat{y}_{t-1}^+, \hat{\theta}_{t-1}, \hat{\omega}_{t-1}, \\ &\hat{r}_{t-1}^x + (\tilde{x}_t - \hat{x}_{t-1}^+), \hat{r}_{t-1}^y + (\tilde{y}_t - \hat{y}_{t-1}^+), \hat{k}_{t-1}]^T\end{aligned} \tag{4}$$

which is compared to the detected features in the examined frame to estimate the final state of the feature.

A feature is declared lost if the verification of $\hat{s}_t^h$ has failed for 5 consecutive frames in order to eliminate false positives during tracking as they would provide inaccurate information about the underlying tissue motion. Failure of the tracking process is determined when 35% of the features have been lost. A low threshold for signalling tracking failure might lead to over-segmentation of the video. However, in the proposed framework it is desirable to acquire motion information from the entire scene without neglecting areas where features might have been lost.

A new episode is defined by re-initializing tracking and automatically selecting a subset of the detected features to track. Non-maximal suppression is applied to select the most salient ones to enable long tracks along time. Salient points that correspond to specular highlights are excluded. In order to represent every area of the examined environment, it is desirable that the tracked features span the whole field of view. The selected features are compared in terms of the difference in the surrounding image area and the most dissimilar ones are selected to initialize tissue tracking as they correspond to different areas of the scene.

2.2 Episode Representation

In the proposed surgical episode representation, probabilistic motion modeling is used to represent the motion of tracked features. In the ideal case, the distribution of a feature's velocity within an episode should be represented by a Gaussian-like distribution around some point. In practice, the motion of a feature can change dramatically over a period of a few frames due to periodic tissue motion or tool-tissue interaction causing free-form tissue deformation. In order to account for such multiple movements, the motion of each feature is modeled as a mixture of K bivariate Gaussian distributions.

Given the motion vectors $U^k = (u^k, v^k)$, $k = 1 \ldots t_e$ of a tracked feature within an episode t_e frames long, the probability of observing the motion vector U^t is given by:

$$P(U^t) = \sum_{p=1}^{K} \omega_p \cdot g(U^t, \mu_p, \Sigma_p) \tag{5}$$

where ω_p is an estimate of the weight of the p^{th} Gaussian in the mixture, μ_p, Σ_p are the mean and the covariance matrix and $g(\cdot, \mu_p, \Sigma_p)$ is the density of the p^{th} component, given by:

$$g(U, \mu_p, \Sigma_p) = \frac{1}{2\pi {\Sigma_p}^{\frac{1}{2}}} \exp\{-\frac{1}{2}(U - \mu_p)^T {\Sigma_p}^{-1}(U - \mu_p)\} \tag{6}$$

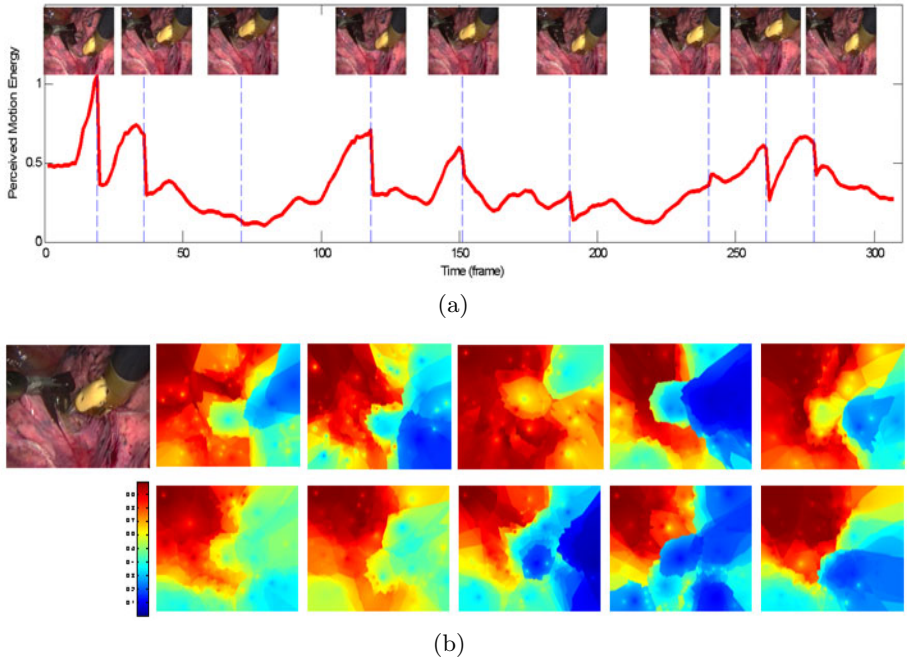


Fig. 1. (a) PME sequence with blue vertical lines indicating episode borders and the first frame of each episode shown. (b) Surgical episode content representation maps.

The number of mixture components is empirically determined by the desired computational complexity and usually ranges from 3 to 5 [9]. In this work, five components have been used. The parameters (μ_p, Σ_p) of the distributions as well as the weights ω_p are learned using the Expectation Maximization (EM) algorithm [10]. For a compact motion model representation, the mean and the covariance matrix of the Gaussian mixture are used and they are estimated as:

$$\begin{gathered} M = \sum_{p=1}^{K}(\omega_p \mu_p) \\ \Sigma = \sum_{p=1}^{K} \omega_p(\Sigma_p + \mu_p \mu_p^T) - (\sum_{p=1}^{K} \omega_p \mu_p))(\sum_{p=1}^{K} \omega_p \mu_p))^T \end{gathered} \tag{7}$$

To this end, the motion models of tracked features are blended to estimate the motion at every point in the scene. Considering a set of N neighboring tracked features, the motion model at a scene point i is estimated as in Eq. (7). The weight of the j^{th} neighbor in the motion blend is estimated as $\omega_j = d_{i,j}^{-1} / \sum_{j=1}^{N} d_{i,j}^{-1}$, where $d_{i,j}$ is the Euclidean distance between the scene point and its j^{th} neighbor.

3 Results

In order to assess the practical value of the proposed framework, quantitative evaluation has been performed on four *in vivo* data sets from robotically assisted

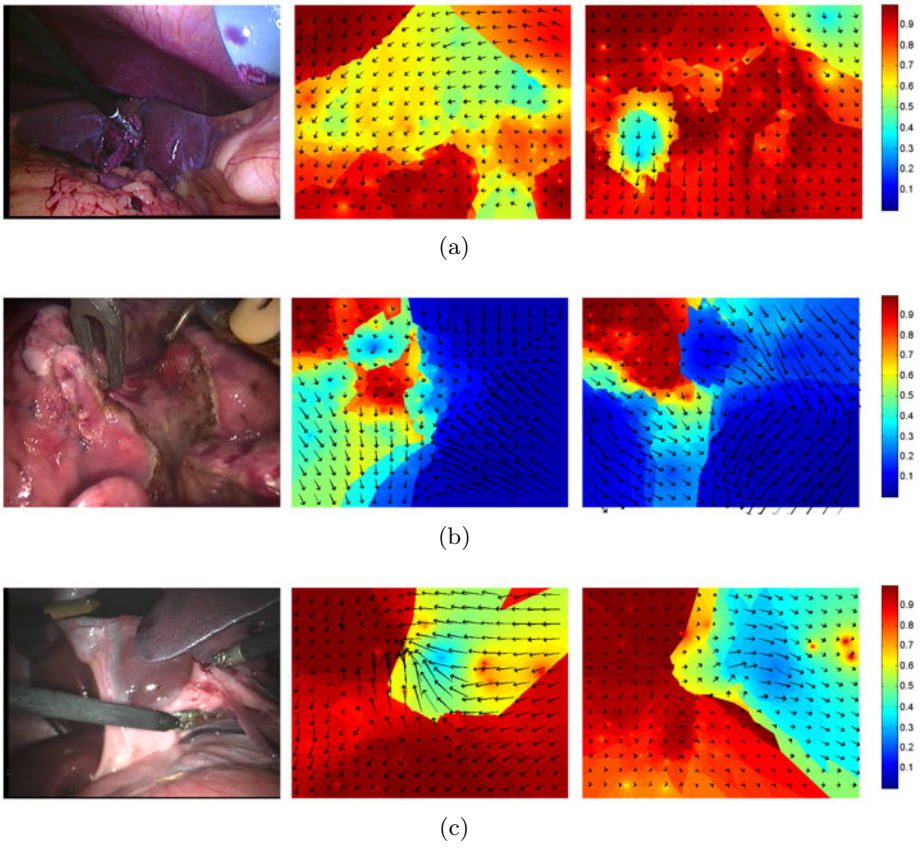
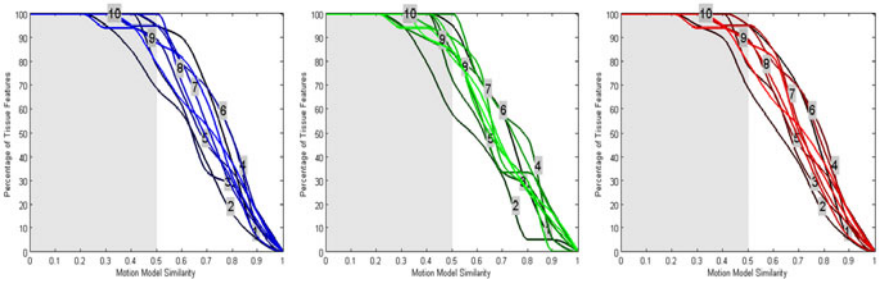


Fig. 2. Motion vectors superimposed on surgical episode content representation maps for MIS sequences with (a) varying camera motions (b) instrument-tissue interaction (c) respiratory motion

MIS procedures involving a variety of representative surgical scenarios such as varying camera motions, occlusion due to the presence of tools and significant tissue deformation. The images are of resolution 360×288 pixels, in line with the output resolution of the available endoscopic tools used in MIS. Visual content information is provided by motion patterns generated by tracking 100 affine-invariant anisotropic regions.

Due to a lack of accepted benchmarking or ground truth for episode segmentation algorithms, it is not possible to perform objective evaluation of the proposed episode segmentation approach, particularly for *in vivo* patient studies. Hence subjectively defined feature tracking is used for performance evaluation. The Perceived Motion Energy (PME) model [11] has been widely used in video segmentation and it is used here to demonstrate the motion activity along surgical episodes and verify that each extracted episode describes a distinct surgical action. The PME model combines the motion intensity and the motion direction of



(a) (b) (c)

Fig. 3. Motion model similarity assessment results for (a)$K = 5$, $N = 5$ (b)$K = 5$, $N = 3$ (c)$K = 3$, $N = 5$. The horizontal axis corresponds to the motion model similarity and the vertical axis to the percentage of examined features on the ground truth.

the tracked features. The PME at frame t of a surgical episode is mathematically defined as:

$$PME(t) = \left(\frac{1}{F(2T+1)} \sum_{i=1}^{F} \sum_{k=t-T}^{t+T} \sqrt{(u_i^k)^2 + (v_i^k)^2} \right) \cdot \left(\frac{max_b(H(t,b))}{\sum_{b=1}^{n} H(t,b)} \right) \quad (8)$$

The first term in Eq. (8) corresponds to the average velocity of tracked features within the frame interval $[t-T, t+T]$ of the episode. The second term expresses the percentage of dominant motion direction within the episode and $H(t,b)$ is the n-bin histogram of the motion vector angles of the tracked features for the interval $[t, t+T]$. The parameter T determines the amount of activity detail captured in the PME and in this work is set to 25% of the episode length.

To validate the proposed episode content representation approach, ground truth data is estimated by manually identifying corresponding features between the first sequence frame and subsequent frames. In order to estimate feature velocities within a surgical episode the ground truth is obtained at equally spaced pairs of consecutive frames that correspond to 20% of the episode length and to a minimum number of 20 frames. In order to reduce the computational complexity of the performance evaluation, 20 of the initially detected features are manually tracked along time to obtain their ground truth motion model which is compared to the one estimated by the proposed approach. The similarity between two motion models i, j is estimated using Matusita's measure [12] which expresses the difference between the covariance matrices and the distance between the means of multivariate distributions, defined in the bivariate case as:

$$S_{i,j} = \frac{2|\Sigma_i|^{\frac{1}{4}}|\Sigma_j|^{\frac{1}{4}}}{|\Sigma_i + \Sigma_j|^{\frac{1}{4}}} \exp\{-\frac{1}{4}(\mu_i - \mu_j)^T(\Sigma_i + \Sigma_j)^{-1}(\mu_i - \mu_j)\} \quad (9)$$

Detailed workflow analysis of an *in vivo* sequence with tissue motion due to respiration and significant instrument-tissue interaction is presented in Fig. 1. The PME pattern in Fig. 1(a) verifies that the extracted surgical episodes describe distinct tool actions within the workflow. Also, the PME level is proportional

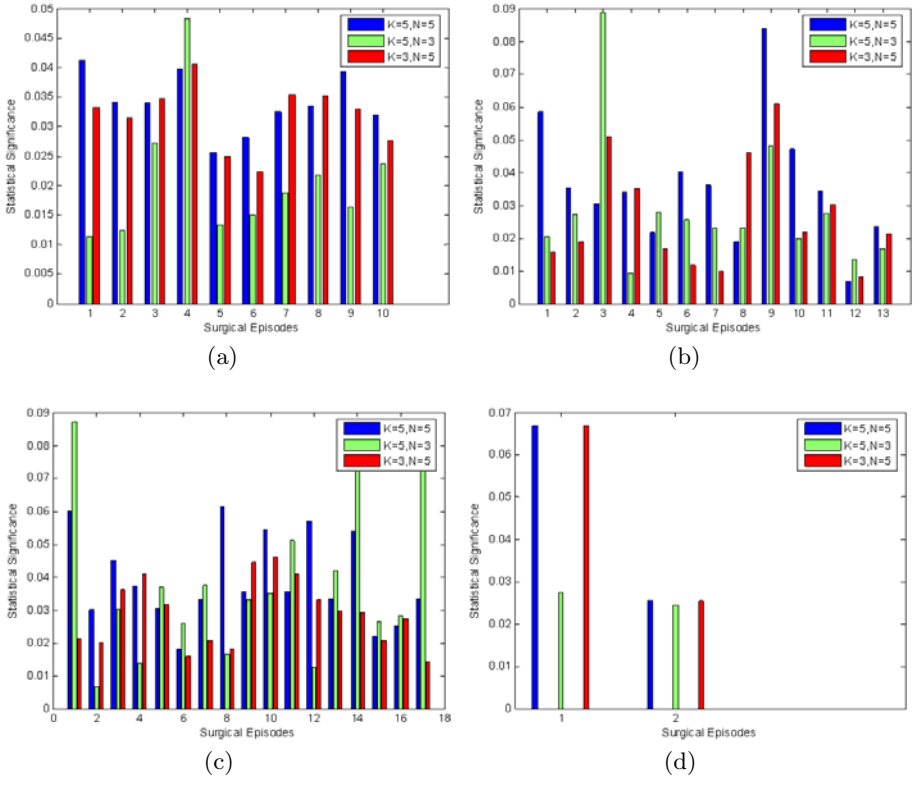


Fig. 4. Statistical significance results for MIS sequences shown in (a) Fig.1 (b) Fig. 2a (c) Fig. 2b (d) Fig. 2c

to the degree of deformation and episodes with no instrument-tissue interaction are characterized by low PME which is mainly due to respiration motion. Observing the PME pattern around the borders of the episodes extracted by the proposed method, it becomes clear that video segmentation based on the PME model (episode borders identified at the peaks of the PME pattern) has missed surgical actions occurring at a small part of the scene (for instance the 3^{th} and the 6^{th} episode borders). This demonstrates the success of the proposed method and suggests that PME is a useful measure for representing overall scene changes but not sufficient enough to detail fine granularity of the movement patterns.

The inherent surgical episodes of the above video sequence are represented in Fig. 1(b). Areas of coherent motion within the episode are graphically classified using a colormap to demonstrate the similarity between the motion of scene points and a reference point (e.g the upper left corner of the scene) estimated using Eq. (9). The colormap representations clearly illustrate instrument induced tissue deformation. The content of a sample of surgical episodes extracted from the rest of the examined MIS data sets is presented in Fig. 2. Arrows have been used to illustrate the motion of discrete scene areas within the episode,

demonstrating global camera motion (a), instrument interaction (b) and respiratory motion (c).

Fig. 3 demonstrates the accuracy of the estimated motion model for each surgical episode of the sequence in Fig. 1 when compared to the ground truth. Each curve point (a, b) corresponds to the percentage b of the examined features on the ground truth with similarity between the ground truth and the estimated motion model higher than a. Fig. 3(a) shows the motion model accuracy when using $K = 5$ mixture components and blending $N = 5$ features. The similarity curves show that the motion of a significant percentage of the scene conforms to the manually defined ground truth. The effect of the number N of neighbors used in the blending process is investigated in Fig. 3(b) by setting $N = 3$. The similarity curves in Fig. 3(c) show the accuracy of motion modeling when 3 mixture components are used.

The ideal similarity curve would be a straight horizontal line at 100%. However, this is not always achievable and the slope of the curve is an indication of the robustness of the proposed episode representation. Since the difference in the performance of the above three parameterization schemes is not distinctive enough, statistical analysis of the similarity scores is performed. The statistical significance of the different parameterizations is evaluated as:

$$mean(S_{norm}^{E_i}) + std(S_{norm}^{E_i}) \tag{10}$$

where, $S_{norm}^{E_i} = mean(S_t^{E_i} - \frac{1}{M}\sum_{j=1}^{M} S_t^{E_j})$ stands for the normalized similarity scores using parameterization E_i, $S_t^{E_i}$ denotes the similarity score at episode t when using parameterization E_i and M is the total number of compared parameterizations. As shown in Fig. 4, the selected parameterization with $K = 5$ and $N = 5$ gives the highest statistical significance in the majority of episodes for all of the examined sequences. The number of mixture components does not affect the accuracy of the proposed representation significantly, while the distance of the estimated representation from the ground truth increases when fewer neighbors are blended.

4 Conclusions

In this paper, we have proposed a novel framework for the succinct representation of the content of MIS data. Probabilistic tissue tracking is used to generate motion patterns that represent visual content complexity and guide the identification of surgical episodes. Surgical episodes are represented by modeling the tissue motion using probabilistic models and blending the models together to acquire information about the motion of the entire scene. The proposed compact, content-based data representation will facilitate surgical workflow analysis and understanding. It identifies episodes of distinct surgical actions providing the required information for their fast and robust reconstruction with techniques such as SLAM or structure from motion. The proposed framework also simplifies the problem of scene reconstruction as the information for feature extraction and data association is already conveyed in the proposed data representation.

References

1. Varadarajan, B., Reiley, C., Lin, H., Khudanpur, S., Hager, G.: Data-derived models for segmentation with application to surgical assessment and training. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 426–434. Springer, Heidelberg (2009)
2. Sikora, T.: The MPEG-7 visual standard for content description-an overview. IEEE Transactions on Circuits and Systems for Video Technology 11(6), 696–702 (2001)
3. Smoliar, S.W., Zhang, H.: Content based video indexing and retrieval. IEEE Multimedia 1(2), 62–72 (1994)
4. Lo, B., Darzi, A., Yang, G.Z.: Episode classification for the analysis of tissue/instrument interaction with multiple visual cues. In: Ellis, R.E., Peters, T.M. (eds.) MICCAI 2003. LNCS, vol. 2878, pp. 230–237. Springer, Heidelberg (2003)
5. Rosen, J., Solazzo, M., Hannaford, B., Sinanan, M.: Task decomposition of laparoscopic surgery for objective evaluation of surgical residents' learning curve using hidden markov model. Computer Aided Surgery 7(1), 49–61 (2002)
6. Ahmadi, S.A., Sielhorst, T., Stauder, R., Horn, M., Feussner, H., Navab, N.: Recovery of surgical workflow without explicit models. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 420–428. Springer, Heidelberg (2006)
7. James, A., Vieira, D., Lo, B., Darzi, A., Yang, G.Z.: Eye-gaze driven surgical workflow segmentation. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 110–117. Springer, Heidelberg (2007)
8. Giannarou, S., Visentini-Scarzanella, M., Yang, G.Z.: Affine-invariant anisotropic detector for soft tissue tracking in minimally invasive surgery. In: IEEE International Symposium on Biomedical Imaging, pp. 1059–1062 (2009)
9. Stauffer, C., Grimson, W.: Learning patterns of activity using real-time tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(8), 747–757 (2000)
10. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the em algorithm. Journal of the Royale Statistical Society, Series B 39(1), 1–38 (1977)
11. Liu, T., Zhang, H.J., Qi, F.: A novel video key-frame-extraction algorithm based on perceived motion energy model. IEEE Transactions on Circuits and Systems for Video Technology 13(10), 1006–1013 (2003)
12. Matusita, K.: Decision rules based on distance for problems of fit, two samples and estimation. Annals of Mathematical Statistics 26, 631–641 (1955)

Improved Precision in the Measurement of Longitudinal Global and Regional Volumetric Changes via a Novel MRI Gradient Distortion Characterization and Correction Technique

Vladimir S. Fonov[1], Andrew Janke[2], Zografos Caramanos[1], Douglas L. Arnold[1], Sridar Narayanan[1], G. Bruce Pike[1], and D. Louis Collins[1]

[1] McConnell Brain Imaging Center, Montreal Neurological Institute, McGill University Montreal Canada
vladimir.fonov@gmail.com, aki@mrs.mni.mcgill.ca, doug@mrs.mni.mcgill.ca, sridar@mrs.mni.mcgill.ca, bruce.pike@mcgill.ca, louis.collins@mcgill.ca

[2] Department of Geriatric Medicine, College of Medicine and Health Sciences, The Australian National University, Canberra Australia
a.janke@gmail.com

Abstract. Reducing measurement variability in MRI-based morphometric analysis of human brain structures will increase statistical power to detect changes between groups and longitudinally over time in individual subjects. One source of measurement error in anatomical MR is magnetic field gradient-induced geometric distortion. This work proposes a method to characterize and compensate for these distortions using a novel image processing technique relying on the image acquisition of a phantom with known geometrical dimensions, without the need to acquire the magnetic field mapping. The method is not specific to any particular shape of the phantom, as long as it provides enough coverage of the volume of interest and enough structure to densely sample the distortion field. The distortions are expressed in terms of spherical harmonic functions, which are then used to define the distortion correction field for the volume of interest. Accuracy of the distortion measurement was evaluated using numerical simulation and reproducibility was estimated using multiple scans of the phantom in the same scanner. Finally, scan-rescan experiments with nine healthy subjects demonstrated that 90% of the distortion (in terms of local volume change) can be corrected with this technique.

Keywords: Geometric distortion, phantom, spherical harmonics, MRI, morphometric analysis.

1 Introduction

Magnetic resonance imaging (MRI) is widely used in many longitudinal studies of normal brain development and aging as well as the evaluation of neurological disease.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 324–333, 2010.

Quantitative measurement of regional anatomical volumes is often used in these studies. Magnetic field gradient nonlinearities result in geometric distortion of the images and this can have a significant impact on the accuracy of volume measures and estimation of volume change between scans [1]. To correct for these distortions, two classes of methods have been described:

1. indirect mapping of the deformation field using a physical phantom with easily identifiable, known structures [2], [3]. MR images of the phantom are used to estimate the deformation field and
2. direct measurement of each magnetic field gradient using specialized hardware. The measurements are then fit to a mathematical model of the gradients, and the inverse problem is solved to estimate a distortion correction field [1], [4].

While the latter direct method is attractive on theoretical grounds, it requires spherical harmonic information specific to each scanner and gradient set to which it will be applied, which in turn requires explicit mapping of the magnetic field strength within the bore of the MRI magnet. Given that these measurements are complex and require specialized equipment, this method generally is only feasible if the spherical harmonics information can be obtained from the scanner manufacturer. In practice, this information is difficult to obtain as it depends on sensitive measurements acquired by medical physicists or engineers.

In the present study, we propose a procedure for measuring and correcting gradient-induced distortions from the first class of indirect methods that uses a simple geometric phantom that is both inexpensive and easy to reproduce. It does not require changing the acquisition protocol; in fact the protocol used for *subject* data acquisition should be used to acquire the necessary data of the phantom for distortion correction. Furthermore, accurate positioning of the phantom at the magnetic center of the scanner is not needed – the phantom simply has to cover the field of view where subject data will be acquired. The distortion correction field is expressed using a spherical harmonic expansion and is based on a mathematical model of the magnetic field gradient [5]. However, unlike the second class of methods described in [4] and in [1], the proposed method does not require mapping of the magnetic field.

We have already reported the effect of the gradient-induced distortions on the longitudinal measurements of whole-brain atrophy [6]. In the current work we expand the proposed technique to compensate for the gradient-induced distortions and demonstrate its effect on the precision of the longitudinal volumetric analysis.

2 Methods

2.1 Distortion Model

Our approach is based on the idea that the geometric distortion field within the scanner may be expressed using a spherical harmonic expansion (SPH). However, in our method we do not try to measure the strength of the magnetic field explicitly to model the distortion. Instead, we use apparent displacements of corresponding points to calculate the coefficients of the coordinate mapping functions (Eq. 1,2):

$$F_{V,n,m}(r,\theta,\phi) = r^n \left[a_{V,n,m}\cos(m\phi) + b_{V,n,m}\sin(m\phi)\right] P_{(n,m)}(\cos\theta) \tag{1}$$

where $V= x,y,z$ are Cartesian coordinates; r, θ, ϕ are spherical coordinates in an "*ideal*"' coordinate system, and $P_{(n,m)}$ are the associated Legendre polynomials, and

$$X_V^* = \sum_{n,m} F_{V,n,m}(r,\theta,\phi) \tag{2}$$

where X_V* are coordinates in the scanner specific coordinate system.

We use least squares approximation to calculate the SPH coefficients based on the dense field of displacements that is obtained by matching the "ideal" representation of the phantom to the acquired MRI scan using iterative nonlinear image registration technique described below. Because we are using a relatively small number of coefficients (105 for the 5th order spherical components expansion), this method is quite robust to random noise or partially missing information. Additionally, for some acquisition protocols it is possible to assume cylindrical symmetry around the Z axis, which decreases the number of coefficients for the 5th order expansion from 105 to 40 yielding an even more robust solution.

2.2 Geometric Phantom

The proposed method does not require any particular phantom to be used in the imaging, as long as it covers the volume of interest (VOI), and contains enough structural features to densely sample the distortion field. It is possible, for example, to use the ADNI phantom [7] for this analysis, or any other phantom with known geometrical information.

For this study we have constructed a phantom consisting of 125 Lego DUPLO® bricks made of ABS plastic (Billund, Denmark) assembled inside a polycarbonate Nalgene® 8L container and filled with a water solution of 0.15mM/L $MnCL_2$ and 2.8g/L NaCL, according to [8] (see Fig. 1). Lego DUPLO® 2x4 bricks were chosen for the construction in order to have a phantom that can easily be reproduced across multiple sites with minimal cost and with high degree of accuracy, as manufacturing tolerances for LEGO bricks are 2μm [9].

2.3 Data Processing

As a preprocessing step, the "ideal" phantom volume is created numerically using the knowledge of the location of the Lego® DUPLO bricks and their geometric properties. A similar procedure is employed for the ADNI phantom.

As a first step of data processing, we invert the intensities of the MRI scan and remove the background using the morphological operators. A rigid body registration is then used to align the "ideal" phantom volume to the acquired scan data (see Fig. 2 A, B, C). Next, the hierarchical iterative process of estimating the coefficients for the geometric distortion correction is run:

1. Start with coefficients describing the identity transformation.
2. For each pair of scanned and "ideal" phantom volumes, calculate a nonlinear mapping [10]. The algorithm parameters (level of blurring and node spacing) are reduced at each iteration in hierarchical fashion, starting with 8mm steps between nodes and 4mm full-width half-max (FWHM) Gaussian blurring and ending with 2mm steps and 1mm FWHM Gaussian blurring.

3. From the dense field of deformation vectors that defines the non-linear mapping recovered in step 2, calculate the coefficients of the SPH using least squares approximation. Note that data from multiple acquisitions of the phantom can be combined to improve the accuracy or enlarge the coverage
4. Calculate a dense vector field for the next iteration using SPH expansion.
5. Repeat the process, starting from step 2 until the last step in the hierarchical schedule.

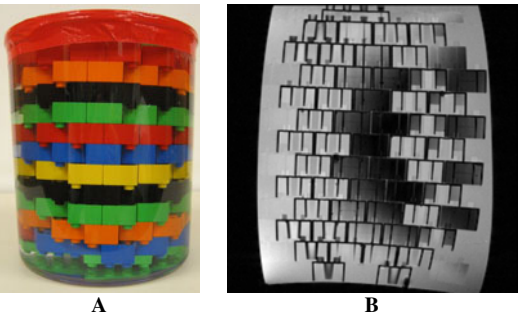


Fig. 1. LEGO DUPLO® phantom, (A) photo of the exterior of the phantom, (B) sagittal slice of the T1w FLASH acquisition (Siemens Sonata 1.5T scanner)

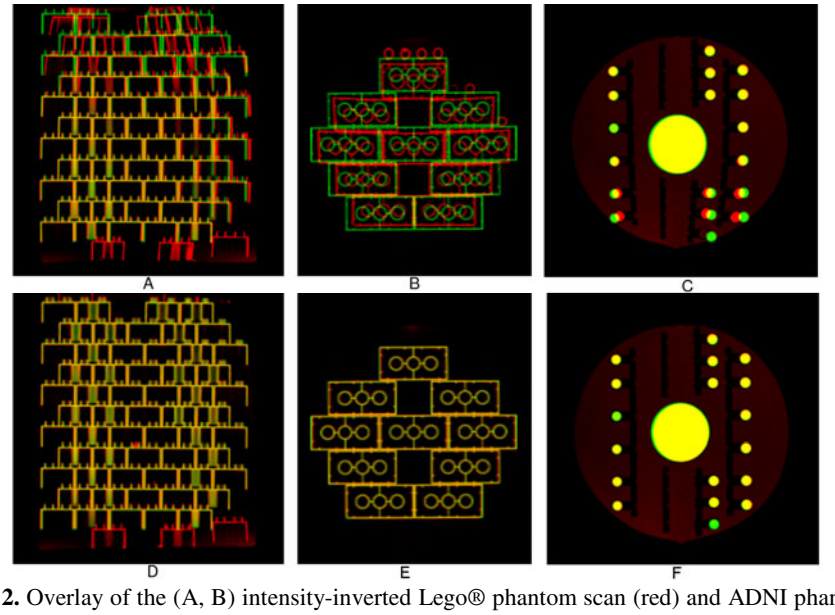


Fig. 2. Overlay of the (A, B) intensity-inverted Lego® phantom scan (red) and ADNI phantom scan (C) and the "ideal" reconstruction (green), before distortion correction and (D,E,F) after distortion correction. Note that where red and green lines are visible, the data from the scanned and ideal phantoms do not align. Yellow structures are well aligned (yellow = red + green). The transverse slice in B and E corresponds to the second row of bricks from the top of the phantom, as seen in the sagittal images in A and D.

Once the process is finished, a distortion correction field is produced that can be applied directly to unwarp scans to compensate for gradient-induced distortions, See Fig. 2-D & E for the result of applying the distortion correction field to the Lego® phantom itself and Fig. 2-F for the results of application of the method to the ADNI phantom scans.

3 Experiments

Three sets of experiments were conducted to validate the algorithm. All were performed with the Lego® phantom. The first used numerical simulations to test the algorithm in a fully controlled environment. The second used two sets of MRI experiments with multiple scans of the phantom to evaluate the precision of the algorithm using multiple acquisitions in the same scanner. Finally, the third set of experiments evaluated how the distortion correction procedure affects volumetric measurements using MRI of human subjects.

3.1 Numerical Simulations

The performance of the algorithm was first characterized with numerical simulations to recover a realistic known geometric distortion (which were recovered from the real experiment). One thousand two hundred numerical simulations were performed. For each simulation, the same receive coil inhomogeneity was applied to the "ideal" reconstruction of the phantom together with small random rotations (-5° - 5° around x, y and z axis), shifts (-5mm - +5 mm in x, y and z directions) and known distortion field. Also, independent instantiations of Rician noise at a SNR of 20 (300 simulations) and 40 (300 simulations) were applied. For each level of SNR, both cylindrically-symmetric and non-symmetric models were applied using a 5^{th} order approximation yielding a total of 1200 simulations. RMS differences between simulated and estimated distortions were computed within a 200 mm diameter sphere centered at the isocenter of the gradients.

3.2 Precision of Distortion Measurements

Precision of the distortion-correction field was estimated by scanning the phantom with a T1-weighted FLASH sequence on a Siemens Sonata 1.5T scanner at different positions (at the isocenter of the magnet, ±10-mm displacement along the X- and Y-axes, and ±50-mm displacement along the Z-axis). Two data processing scenarios were used, both using a 5^{th}-order symmetric model.

In the first scenario, each scan was processed independently and RMS differences between distortion fields recovered from these seven scans were measured within a 200 mm diameter sphere centered at the magnet's isocenter.

In the second scenario, a leave-one-out approach was used. RMS differences between recovered distortion fields (estimated on the remaining 6 scans) were measured within a 200 mm diameter sphere centered at the magnet's isocenter.

3.3 Human Experiments

Nine healthy subjects were scanned three times on the same day on a Siemens Sonata 1.5T scanner using the same sequence (sagittal T1w 3D FLASH, TE=10 ms, TR=22 ms, Flip angle=30°, voxel size 1x1x1.5mm, matrix size 256x256x144). The study was approved by the Research Ethics Board of the Montreal Neurological Institute, and informed consent was obtained from all participants. Each subject was first scanned at the isocenter of the gradients (termed baseline scan, *S*). Then, the scanner bed was shifted forward 50-mm and the scan repeated (Z-shifted scan, *Z*). The subject was then taken out of the scanner, repositioned at the isocenter and scanned again (reposition scan, *R*). The distortion correction field was estimated using three phantom scans acquired with the same sequence and parameters: one at the magnet isocenter, one shifted by +50 mm in the bore (Z) direction and the third, -50 mm in the bore direction.

Each of the subject images were corrected for intensity non-uniformity using the N3 algorithm [11]. The image pairs (*S*-*R* and *S*-*Z*) were co-registered in native space using a rigid-body transformation using a cross-correlation objective function; both pairs of scans were resampled in a half-way space. Non-linear registration [10] was used to estimate the residual anatomical misalignment due to the geometric distortion inside the brain. Local volume differences due to geometric distortion were estimated by computing the Jacobian-determinant of the non-linear deformation field. Finally, the Jacobian determinant field was resampled into stereotactic space using linear registration of the scan into MNI152 space [12].

The procedure described above was applied to the subject data with and without distortion correction to evaluate the improvement due to the distortion-correction method. This was evaluated by performing statistical analysis on smoothed (Gaussian kernel with FWHM of 10-mm) Jacobian-determinant fields in a similar manner to that described in [13]. Absolute Jacobian-determinants were averaged between subjects in a voxel-wise manner and two-sided *t*-tests were performed to identify voxels where absolute distortion was significantly smaller (or greater) after distortion correction.

4 Results

4.1 Numerical Simulations

The numerical simulations demonstrated that, without the assumption of cylindrical symmetry, the RMS difference between known and recovered distortion correction fields was ~0.6 mm for SNRs of 20 and 40; with the assumption of cylindrical symmetry, this difference decreased to ~0.3 mm (see Fig. 3). The difference between cylindrical and non-cylindrical models is statistically significant with $p<0.001$.

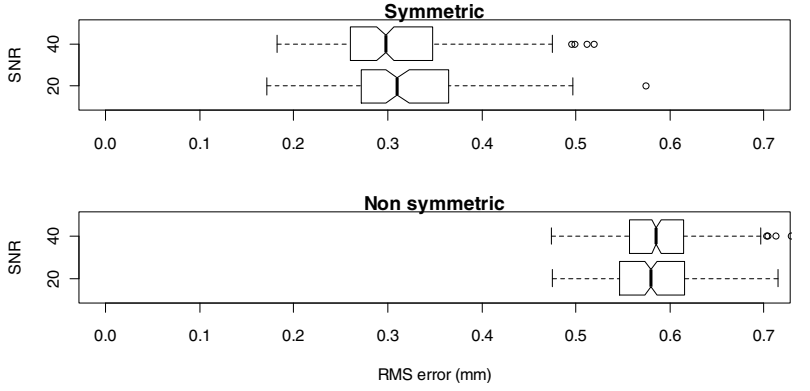


Fig. 3. Numerical simulation, RMS difference between known and recovered distortion for SNR of 20 and 40. Cylindrically symmetric (top) and non symmetric (bottom) model of distortion field.

4.2 Reproducibility of Distortion Measurements

The repeated acquisitions of the Lego® phantom demonstrated that the distortion field could be reproduced between phantom acquisitions with an RMS difference of 0.4 mm across the seven trials in the case where a single acquisition was used to estimate the distortion. When 6 of 7 acquisitions were used to estimate the distortion field, the RMS difference between the 7 estimates was 0.04 mm.

4.3 Human Data

Fig. 4 shows intensity difference maps between baseline and Z-shifted scan of one subject. Qualitatively it is visible that uncorrected data shows much greater variability in the cortex. Over all subjects, voxel-wise analysis shows that volumetric distortions within more than 86% of the brain volume were reduced in a statistically significant fashion. Fig. 5 shows the average Jacobian-determinant map of the distortions before correction and Fig. 6 - after correction (i.e the residual error), where a value of 0.1

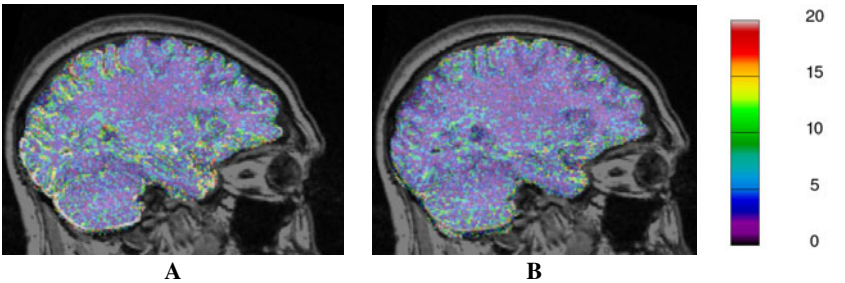


Fig. 4. Intensity difference images (scans *S* - *Z*) before (A) and after (B) distortion correction. Note the reduced intensity difference after correction, especially near the edges of sulci.

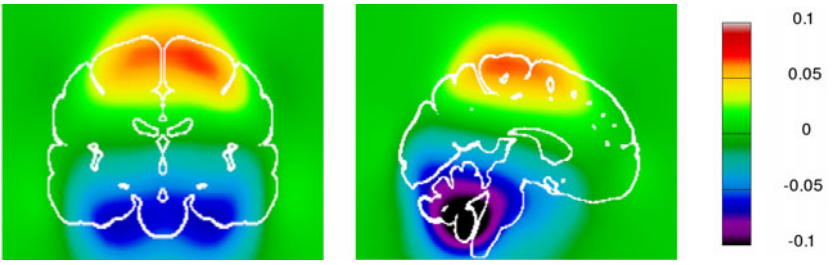


Fig. 5. Average Jacobian-determinant map, uncorrected images, comparing scans S & Z. Shifting the subject by 50mm along the bore of the magnet results in an apparent increase in volume of 7-8% at the vertex and a decrease of approximately 10% in the region of the cerebellum.

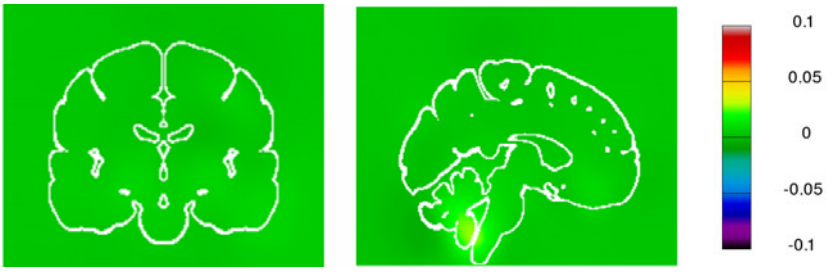


Fig. 6. Average Jacobian determinant map comparing *S* & *Z* images, after applying the distortion correction (residual error). Note almost perfect correction throughout the field of view.

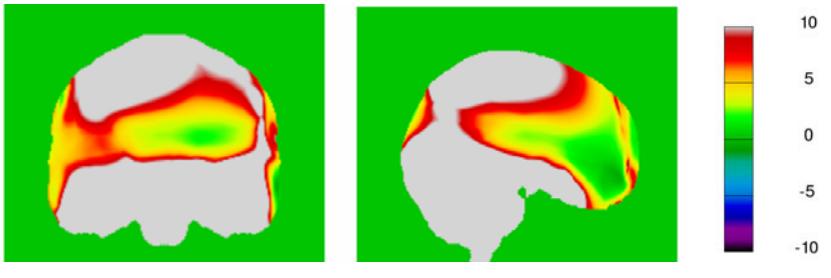


Fig.7. Statistical t-map of the improvement (positive t means that absolute local-volume-variation is smaller after distortion correction). Note that the large white region corresponds significant improvement, ie. t>10.

corresponds to a 10% change in local volume. Fig. 7 shows the *t*-map of the statistical significance of the reduction of average absolute Jacobian-determinants after distortion-correction: with the t-threshold set at 3.0 (for a false-discovery rate of 0.01), *(i)* 86% of the brain volume showed significantly decreased distortions (*i.e.*, was associated with a *t*-value > 3.0); but *(ii)* only 0.1% of the brain volume showed increased distortions (*i.e.*, was associated with a *t*-value < -3.0). On average, the mean (between-subjects) absolute Jacobian-determinant within the brain was reduced by 90%.

5 Discussion and Conclusions

Gradient-induced geometric distortion can be an important component of the variance of estimated structure volumes. Indeed, the results shown in Fig. 5 demonstrate that a 5cm shift in along the bore of the magnet can result in large apparent changes in volume: regions near the top of the brain increase in volume by 7-8% while structures near the cerebellum shrink by approximately 10%. While a 5cm shift might appear extraordinarily large, a previous study of clinical trial data [6] has shown that the repositioning error in the z direction can be quite large and sometimes larger that 10cm. Even though most data is acquired within 35mm, the subjects' brains will be submitted to different parts of the geometric distortion field and this will result in unwanted variability in volume estimation. Given that many studies aim to detect (very) small changes, for example 1%/year change in hippocampal volume in patients with mild cognitive impairment, any geometric distortion will result in added measurement variance and in a reduced power to detect change.

This paper has demonstrated a phantom-based technique for distortion correction that is both simple and inexpensive to use. In our scanner, this procedure reduces the distortions significantly for 86% of the volume covered by the brain, reducing the magnitude of the distortions by 90% (in terms of local volume change). In the remaining 14% of the brain, geometric distortion was negligible before correction. The practical consequence of application of such a procedure will be reduced measurement variability and improved power to detect change.

The distortion correction method presented here is shown to be reproducible to within 0.4 mm in terms of the distortion recovery in the case of one phantom acquisition, and as good as 0.04 mm when multiple phantom acquisitions are combined. This is partly due to the fact that, by combining multiple acquisitions we are simply increasing number of averages, thus increasing SNR. In addition, by combining scans collected at different locations within the scanner we are increasing the volume within which our method works as interpolation rather then extrapolation, thus reducing uncertainty in defining the parameters of the distortion field. Restricting the model of the distortion field to be cylindrically symmetric also improves the reproducibility by reducing the number of unknowns, but special care should be taken in ensuring that this assumption is applicable to the particular acquisition protocol: for example if a read out direction other than the z-direction is used this assumption becomes invalid and the full 3D model should be used to model the deformation field.

Finally, it is important to note that the actual magnitude of the distortions estimated here are specific to the scanner used in this study (Siemens Sonata 1.5 T). Nevertheless, whereas the degree and pattern of distortion will vary between scanner manufacturer and model, the procedure is applicable to all scanners. While a few of the latest-generation scanners have manufacturer-supplied 3D distortion-correction capabilities, the majority of scanners currently in use do not, or have limited 2D distortion correction capabilities.

It also is worth noting that this method may be applied retrospectively to the studies in which a geometrical phantom (for example, the ADNI phantom) was acquired, provided that the geometrical description of the phantom used is available or may be measured.

Instructions for the Lego® phantom creation and the analysis software are available at http://en.wikibooks.org/wiki/MINC/Tools/DistortionCorrection . Application of this method improves volumetric scan-to-scan reproducibility, and could be adapted to the requirements of a specific study and available geometrical phantom, and thus will have a positive effect on the power of longitudinal and cross sectional studies

References

1. Jovicich, J., Czanner, S., Greve, D., Haley, E., van der Kouwe, A., Gollub, R., Kennedy, D., Schmitt, F., Brown, G., Macfall, J., Fischl, B., Dale, A.: Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. Neuroimage 30, 436–443 (2006)
2. Wang, D., Doddrell, D.M., Cowin, G.: A novel phantom and method for comprehensive 3-dimensional measurement and correction of geometric distortion in magnetic resonance imaging. Magnetic Resonance Imaging 22, 529–542 (2004)
3. Baldwin, L., Wachowicz, K., Thomas, S., Rivest, R., Fallone, G.: Characterization, prediction, and correction of geometric distortion in 3 T MR images. Medical Physics 34, 388–399 (2007)
4. Janke, A., Zhao, H., Cowin, G.J., Galloway, G.J., Doddrell, D.M.: Use of spherical harmonic deconvolution methods to compensate for nonlinear gradient effects on MRI images. Magnetic Resonance in Medicine 52, 115–122 (2004)
5. Krieg, R.S.O.: Method for three-dimensionally correcting distortions and magnetic resonance apparatus for implementing the method, Patent 6,501,273 (2002)
6. Caramanos, Z., Fonov, V.S., Francis, S.J., Narayanan, S., Pike, G.B., Collins, D.L., Arnold, D.L.: Gradient distortions in MRI: Characterizing and correcting for their effects on SIENA-generated measures of brain volume change. NeuroImage 49, 1601–1611 (2010)
7. Magphan® Quantitative Imaging Phantom (ADNI) (2010)
8. Price, R.R., Axel, L., Morgan, T., Newman, R., Perman, W., Schneiders, N., Selikson, M., Wood, M., Thomas, S.R.: Quality assurance methods and phantoms for magnetic resonance imaging: Report of AAPM nuclear magnetic resonance Task Group No. 1. Medical Physics 17, 287–295 (1990)
9. LEGO Company Profile (2007)
10. Collins, D.L., Evans, A.C.: Animal: Validation and Applications of Nonlinear Registration-Based Segmentation. International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI) 11, 1271–1294 (1997)
11. Sled, J.G., Zijdenbos, A.P., Evans, A.C.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging 17, 87–97 (1998)
12. Collins, D.L., Neelin, P., Peters, T.M., Evans, A.C.: Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. Journal of Computer Assisted Tomography 18, 192–205 (1994)
13. Leow, A.D., Klunder, A.D., Jack, J.C.R., Toga, A.W., Dale, A.M., Bernstein, M.A., Britson, P.J., Gunter, J.L., Ward, C.P., Whitwell, J.L., Borowski, B.J., Fleisher, A.S., Fox, N.C., Harvey, D., Kornak, J., Schuff, N., Studholme, C., Alexander, G.E., Weiner, M.W., Thompson, P.M.: Longitudinal stability of MRI for mapping brain change using tensor-based morphometry. NeuroImage 31, 627–640 (2006)

DVV: Towards a Taxonomy for Mixed Reality Visualization in Image Guided Surgery

Marta Kersten-Oertel[1], Pierre Jannin[1,2], and D. Louis Collins[1]

[1] McConnell Brain Imaging Center, MNI, McGill University, Montreal, Canada
[2] Visages-U746 INRIA, INSERM, CNRS, Universit de Rennes 1, Rennes, France

Abstract. We introduce the DVV taxonomy (**D**ata, **V**isualization processing, **V**iew) which defines each of the major components of a mixed reality image-guided surgery (IGS) system. We propose that these components should be considered and in turn validated for acceptance of a system into the operating room. A taxonomy of IGS visualization systems is a step towards developing a common language that will help developers and end-users discuss and understand the constituents of a visualization system. We demonstrate the usefulness of the taxonomy by classifying fifteen state-of-the-art research papers in the domain of mixed reality visualization IGS systems.

1 Introduction

In image-guided surgery (IGS), preoperative plans, patient models and graphical representations of surgical tools are displayed to guide surgeons in their tasks. The tools and data sets are fused into a mixed reality providing the surgeon with a view beyond the visible anatomical surface of the patient, thereby reducing patient trauma, and potentially improving clinical outcomes.

Mixed reality is the area on the reality–virtuality continuum [11] between reality, the unmodelled real environment, and virtual reality (VR), a purely virtual and modelled environment. The point on the continuum at which an environment lies will correspond to the extent to which the environment is modelled and whether real or virtual objects are introduced into this environment. Upon this continuum lie augmented reality (AR) and augmented virtuality (AV).

In AR the environment is real; it is a physical location in four dimensions and *virtual objects*, which are digital representations and models of real objects, are added to this real environment. In AV, *real objects*, which are unmodelled physical objects, are introduced into a virtual or digital environment.

The objective of this paper is to define a taxonomy of mixed reality visualization systems for IGS in order to (1) introduce a syntax and framework within which mixed reality IGS systems may be discussed, analyzed, and evaluated and (2) allow a better understanding of the most relevant and important components of a mixed reality visualization system. When developing a common language with which to discuss mixed reality IGS systems, past and current systems can more easily be compared and analyzed. Proper analysis and validation of the

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 334–343, 2010.

major constituents of current systems will allow future developers to more easily recognize which components can be reused and which need improvement or new solutions. Such an analysis and evaluation should facilitate a greater presence of future systems in the OR.

In this paper, we first describe the taxonomy in Section 2, then give an example of how the taxonomy was used to analyze 15 state-of-the-art mixed reality IGS visualization systems in Section 3. Conclusions are given in Section 4.

2 Taxonomy for Mixed Visualization in IGS

In order to provide surgeons with a tool to effectively plan out surgeries three main points should be considered: (1) which of the abundance of available data should be used, (2) how it can be effectively merged and visualized, and (3) how it should best be displayed and interacted with. We therefore, classify a mixed

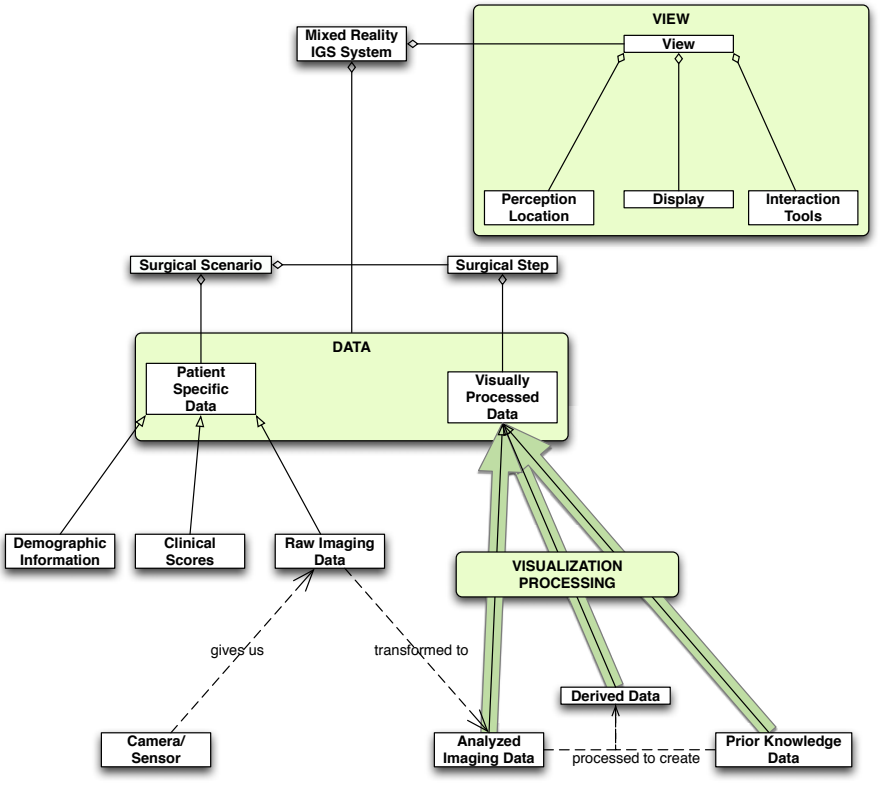


Fig. 1. The three factors of the DVV taxonomy (data, visualization processing and view) are shown in rounded boxes. Subclasses (solid-line arrows) and the relationships between them (dashed-line arrows) are also shown. Diamond arrow heads represent aggregation. The surgical scenario is associated with both visually processed data and view classes.

reality IGS system based on three factors: data, visualization processing, and view (see Fig. 1).

In addition to these three factors, we use the notion of a surgical scenario, which is associated with both data and view classes, to allow us to describe dynamic systems that change based on the end-users' needs during a given task. The surgical scenario allows us to determine (1) what type of visualization data should be shown at a particular point in time of the surgery, (2) where it should be viewed (perception location/display) and (3) how the data may be interacted with at that step in the surgery (interaction tools). The surgical scenario describes the type of surgery and in particular the number of surgical steps that are executed to perform the surgery. Each surgical step describes the action to be done, its associate precision and its completion time.

In the remainder of this section we define the three DVV factors: **D**ata, **V**isualization Processing, and **V**iew and give justification for the use of each factor for classification.

2.1 Data

Two main superclasses of data are considered: patient specific data and visually processed data (with analyzed, derived and prior knowledge subclasses). In general, the different subclasses of data may be directly viewed or may undergo one or more transformations to become visually processed data. Detailed data classes and subclasses are used within the taxonomy for easy specification of all types of data that may be presented to the end-user.

The purpose of using data as part of our classification is to highlight the importance of deciding which of the available data should be shown to the end-user. As the number of possible datasets that can be acquired increases, there is a stronger demand for integration of the different modalities and numerous visually processed data sets into a coherent visualization. It is important that careful consideration be given as to what type of data should be visualized and at what point during the surgery it is useful. The latter is taken care of by considering the surgical scenario.

Patient Specific Data. Patient specific data is particular to the patient; it may include demographics, clinical scores or raw imaging data. Raw imaging data is the direct output of the acquisition system. It is transformed into analyzed imaging data and only then visualized.

Visually Processed Data. Visually processed data is presented to the end-user by means of the view. It may be analyzed data, derived data or prior knowledge data. It is shown within a particular surgical scenario and more specifically for a specific surgical step. Different surgical steps may use a different representation of a given instance of visualization data or a different view. The visually processed data therefore, should show the end-user the most relevant data in its most informative pictorial representation at each point within the surgery.

Analyzed Imaging Data is raw imaging data that has undergone a transformation to become an object. Instances of the analyzed imaging data in IGS are: point, line, plane, contour, surface, wireframe, and volume. As an example, whereas raw imaging data would be the direct output of the magnetic resonance scanner, the corresponding analyzed imaging data could be the slices rendered as a volume.

Prior Knowledge Data is derived from generic models and is available preceding a surgery. Instances of prior knowledge data include: atlases, uncertainty information about the IGS system, surgery road maps, prior measurements, and tool models.

Derived Data is obtained from processing either only patient specific data, for example, uncertainty due to the calibration, registration, and/or segmentation process, or patient specific data and prior knowledge. For example, brain regions may be segmented using an atlas. Instances of derived data include: labels, uncertainty information specific to the patient/type of surgery, and measurements such as tumour volume and distances between regions of interest.

2.2 Visualization Processing

The visualization processing component of the taxonomy represents the specific visualization techniques or transformations on the data that are used to provide the best pictorial representation of the data for a particular task at a given surgical step. Numerous techniques have been developed for visualizing medical data: non-photorealistic rendering, illustrative techniques, photorealistic rendering, colour coding, using transparency, adding depth cues, and using saliency methods (such as highlighting regions of interest).

By choosing an appropriate visualization technique it is possible to increase the diagnostic value of the original data. For this reason we use visualization processing as the second component for classification in the DVV taxonomy.

2.3 View

The view of the system is the end-product of the visualization process, and therefore the part of the system with which the user interacts. It has three major components: the display, the perception location and the interaction tools. The view is used to classify mixed reality IGS systems as it is perhaps the most relevant component to the end-user. It enables a system description in terms of *where* and *how* information is presented to the end-user and how the end-user can interact with the system.

Perception Location. The perception location is the area or part of the environment where we focus our attention in order to benefit from the mixed visualization. The perception location may be the patient, a digital display, a surgical tool, or the real environment. Depending on where the visualization system lies on the reality-virtuality continuum, the perception location may be either a real of virtual environment.

Display. The display is the particular technology that is used to present data to the end-user. Displays fall into one of two classes, those in which images are projected onto a 2-D technology, for example a computer monitor, or those which provide a three-dimensional impression of the scene or object. Those that provide 3-D impressions fall into one of two categories: (1) binocular stereoscopic, systems which require the user to use special head gear or glasses (e.g. stereo glasses and HMDs) and (2) autostereoscopic displays, which do not such as multi-view lenticular displays, videography and holography. The surgical microscope, which allows for both 2-D and 3-D data representation, has also been used to fuse virtual objects with the real scene in IGS.

Interaction tools. Interaction tools are part of the user interface. We suggest two major subclasses: *hardware interaction tools* which are the physical devices that a user employs in order to manipulate data and *virtual interaction tools* which allow the user to manipulate the pose, and view of the data as well as the visualization parameters of the data. Examples of hardware devices for interaction used in IGS include a mouse, a keyboard, a representative tangible object (e.g. surgical tool), a haptic device, or a data-glove. Instances of virtual interaction tools include transfer functions, volume cutting, voxel peeling, clipping planes, turning data visibility on and off, and in general adjusting data and viewing parameters such as colour, brightness, contrast, and transparency.

3 Using the DVV Taxonomy for IGS System Classification: An Example

In order to validate our taxonomy we studied 15 publications chosen at random from a database of 84 publications. Only publications in the area of mixed reality visualization systems for IGS with a well-defined aim of study and precise description of an entire visualization system were chosen. The publications described a system for use in the OR rather than a simulator for diagnosis, planning, or training. A summary table of the 15 randomly chosen articles using the taxonomy components as columns is given in Table 1.

3.1 Data

All of the papers specified the sensor by which the data was acquired (i.e., raw imaging data) and all but one paper and all but one paper ([4]) described the analyzed imaging data and how the analyzed imaging data was visualized. Very few papers, however, mentioned the use of derived or prior knowledge data.

Raw and Analyzed Imaging Data. In the majority of papers, either a surface or a wireframe representation of the analyzed data was used. In the case of systems which used ultrasound data [10,15,16] 2–D slices were used. The use of simple representations such as planes, surfaces, contours and wireframe, allows for real-time refresh rates. Although the use of volume data may be desirable, this is often not feasible due to the requirement of real-time rendering for

Table 1. In the above table we use the DVV taxonomy to classify 15 publications in the domain of mixed reality IGS systems. N/S (not specified) was used when particular components or classes were not discussed within the paper. Symbols used: **A**=Analyzed visually processed data, **D**=Derived visually processed data, **PK**=Prior Knowledge visually processed data.

AR /AV	REF	DATA			VISUALIZATION PROCESSING	VIEW		
		RAW	ANALYZED	VISUALIZABLE		PERCEPTION LOCATION	DISPLAY	INTERACTION
AR	Archip [1]	MRI, fMRI, DTI, intra-op MRI	Plane, surface	**A:** Colour coded transparent surfaces, cross-sectional slices, triangle models illustrate 3D region of activation **D:** N/S **PK:** N/S	*Colour coding* (ROIs), *transparency* (to merge different modalities)	Patient	N/S	N/S
	Bichlmeier [2]	CT or MRI + real-time video	Surface	**A:** Transparent video & surfaces **D:** N/S **PK:** N/S	Video *transparency* according to skin curvature & observer line of sight, *lighting* (Phong shading, virtual shadow), *adding emphasis* (red border around ROI)	Patient	HMD	Red border thickness user specified, observer changes vision channel into patient by moving head
	Birkfellner (Varioscope AR) [3]	CT	Surface or wireframe	**A:** Surface mesh rendered in mono or stereo **D:** N/S **PK:** N/S	*Depth cue* (stereo rendering used), *state visualized* (surface changed to wireframe based on proximity to structure to be localized)	Patient	HMD	Rotations used to get new view
	Blackwell [4]	CT MRI	N/S	N/S	N/S	Patient	Half-silvered mirror	N/S
	Das [5]	CT, US, MRI	Plane, wireframe	**A:** Yellow wireframe cylinder for target lesion overlaid on CT cross-section **D:** N/S **PK:** Needle = blue cylinder with small sphere for tip	*Saliency* (white perimeter around plane), *state visualized* (needle trajectory green where extrapolated)	Patient	HMD	N/S
	Edwards00 (MAGI) [6]	MRI	Surface	**A:** Vessels + tumour rendered as wireframe **D:** N/S **PK:** N/S	*Colour coding* (blue filter for better colour separation between MRI overlay and real image)	Patient	Microscope	N/S
	Glossop [8]	CT	Points, lines	N/A	N/A, laser displays only vector objects	Patient	Laser Projection	N/S
	Hansen [9]	Angiography	Contour	**A**: Silhouette of vessels and tumours **D:** N/S **PK:** N/S	*Illustrative rendering* (distance encoded silhouettes, varying stroke texture, and silhouette thickness, contour lines)	Patient	N/S	N/S
	Sielhorst [14]	CT or MRI	Surface or volume or wireframe	**A:** Bone mesh/surface/ volume and transparent/opaque window with glass/highlighted skin **D:** N/S **PK:** N/S	*Compared visualizable data*: mesh, surface vs. volume and transparent vs. opaque window and glass/highlighted skin	Patient	HMD	Track head movement which moves virtual window
	Stetten [15]	Intra-op US	Plane	**A:** US slice **D:** N/S **PK:** N/S	N/S	Patient	Half-silvered mirror (stand alone or fixed to US probe)	N/S
	Wacker [16]	MRI	Plane, contour	**A:** Lesions represented as discs + MRI slice **D:** N/S **PK:** needle = blue cylinder, with red sphere tip	*State visualized (*needle yellow where extrapolated)	Patient	HMD	N/S
AR & AV	Linte [10]	MR, intra-op US	Plane, surfaces	**A:** Colour coded surface model **D:** N/S **PK:** Coloured surgical instruments, mitral valve and valve-insertion tool.	*Colour coding* (ROIs), *transparency* (to overlay US on cardiac model)	Patient or Monitor	HMD or Monitor	N/S
	Paul [13]	MRI, fMRI/ MEG, camera image from each microscope ocular	Surface, wireframe	**A:** 3D Colour coded surfaces, 3D textured smooth surface mesh **D:** N/S **PK:** N/S	*Colour coding* (pre-op surfaces), *transparency* (to overlay texture mapped intro-op images and pre-op surfaces)	Monitor/Patient	Monitor/ Microscope	Toggle each component of the scene on and off change opacity or colour, change pose of the overall scene
AV	Gering [7]	MRI (intra-op + pre-op), MR angiography, fMRI PET, &SPECT	Plane, surface	**A:** Colour coded structures of interest **D:** Label layer/segmented objects shown as boundary around key structures **PK:** Needle = orange cylinder	*Colour coding* (structures, boundary around key structures, biopsy target), *transparency (*volume overlays of intra-op data)	Monitor	Monitor	Change transparency of structures, change pose, use slider to move through volume slices
	Nicolau [12]	CT + intra-op video	Surface	**A:** Colour coded structures of interest, target 2m green sphere **D:** Distance to target shown numerically, crosshairs show optimal trajectory **PK:** N/S	*Colour coding* (structures), *adding emphasis* (ROI outlined with dotted line on live video)	Monitor	Monitor	Change transparency, visibility and colour of objects, rotate, translate zoom and change point of view of virtual camera

intra-operative surgical decision making and image guidance. The publication by Sielhorst *et al.* [14] which compared volume rendering to surface and wireframe rendering, showed that volume rendered data currently does not allow for as fast and precise interaction as that of surface or wireframe data. As processor

speeds increase allowing for interactive refresh rates, the use of volume data may become more common.

Visually Processed Data. With the exception of three publications which focused on novel visualization techniques [2,9,14], little attention was given to how data should be presented to the user in order to provide the most informative representation and enable a good spatial and structural understanding of the data. The majority of the visually processed data described in the publications were colour coded regions of interest [1,5,7,10,13], transparent data [2,7,14] or wireframe meshes [3,5,6,13].

Four papers ([2,5,7,16]) described prior knowledge data in terms of how tools were visualized. The others did not specify whether or not prior knowledge data was used and how it was visualized. In three of the systems which described onscreen tool depiction [5,7,16], a biopsy needle was represented as a cylinder. In Bichlmeier *et al.'s* AR surgical system [2], surgical tools which are virtually extended into the patient cast a realistic shadow providing appropriate visual feedback for improved navigation.

Only two of the systems in the 15 publications described the visualization of *derived data*. Nicolau *et al.* [12] numerically displayed the distance from the tool held by the end-user to the target anatomy and represented the computed optimized trajectory for the surgical tool as a cross-hair. Gering *et al.* [7], used segmented objects derived from a "label layer" to compute the boundaries that were depicted around key structures on the slices of an MR image.

3.2 Visualization Processing

In the selected publications, visualization processing was limited to colour coding structures [1,7,6,10,12], adding emphasis by outlining regions of interest [2,5] and using transparency to combine modality data [1,2,7,10,13,14]. The use of simple visualization techniques such as colour coding and transparency do not always support understanding of the structure and spatial relationships of the data. The use of transparency alone to merge modality data can complicate the perception of relative depth distances and spatial relationships between surfaces. More sophisticated techniques are needed for better understanding and interaction of the visually processed data.

3.3 View

In all but two papers [1,9] the perception location and display were specified. The interaction component however, was rarely mentioned and in the few publications where it was mentioned the discussion was limited.

Perception Location. In the majority of the papers, the perception location was the patient. This suggests that current research is focusing on developing systems where the surgeon or resident does not have to look away from the

surgical field of view and the surgeon or resident benefits most from the visualization. The use of patient as perception location limits the display device to be either laser projection [8], half-silvered mirrors [4,15] or HMDs [2,5,10,14,16]. In the selected publications only Nicolau *et al.* [12] and Gering *et al.* [7] used an external digital monitor as the sole perception location.

Display. In half of the selected publications HMDs were used as display device. HMDs can be either optical see-through, where half-transparent mirrors are used to reflect computer-generated images on top of the real world to the user, or video see-through, where video images of the real world are captured using two video cameras attached to the head gear and are combined with computer generated images and then viewed by the user. With the exception of Birkfellner *et al.'s* head-mounted optical see-through microscope (Variscope AR) [3], the remainder of the selected publications used video see-through HMDs [2,14,16].

Half-silvered mirrors or tomographic reflection were used in two of the selected publications [4,15]. With this method, a computer generated image is reflected through a semi-transparent or half-silvered mirror such that the generated images are projected onto the patient. In Blackwell *et al.'s* work [4], shutter glasses were used and tracked so that the virtual objects could be seen in the correct position and in stereo. Stetten *et al.* [15] developed a traditional mirror overlay system, and a portable system where a mirror and a mini flat panel display were attached to an ultrasound probe. Their system allowed for image overlay onto the patient without tracking and arbitrary slice views because of the free movement of the ultrasound probe.

HMDs and tomographic reflection, unlike monitors, allow for *in situ* visualization. Computer monitors, however, are also used in state of the art systems [10,13,12]. Advantages of monitors include the ability for multiple users to benefit from the visualization, and the fact that monitors are already available in the OR. The latter implies that there is no need to introduce a new display device making the use of a digital monitor both cost efficient and non-intrusive.

In two of the chosen publications [3,6], a surgical microscope or variation there of were used. The use of a device that is already present and used in the OR environment seems a logical solution for a visualization system as it may reduce the amount of disruption of the workflow of the surgeon and seamlessly fit into the infrastructure of a surgical navigation system. In the selected publications, both the MAGI (microscope-assisted guided interventions) [6] system and the Varioscope AR [3] allow for correct stereoscopic visualization of the virtual features presented in the optical path of the microscope. In the former, this is achieved by displaying an offset stereogram image of the virtual information into each of the microscope oculars, and in the latter by means of two miniature VGA displays, one meant for each eye.

Interaction. The discussion of the virtual interaction component in the selected papers with few exceptions was limited to: allowing the user to rotate to get a new view of the data [3,12], allowing the user to toggle components on and off, change opacity or colour or objects [7,12,13], navigate through the scene, and change the position of the virtual camera [12,13].

The portable tomographic reflection system fixed to an US probe developed by Stetten *et al.* [15] may also be considered as a hardware interaction tool. The probe is operated by the end-user to view an ultrasound slice within the patient, and in order to get different slice views the user simply changes the orientation of the ultrasound probe.

4 Conclusions

In this work, we have described the DVV taxonomy based on Data type, Visualization Processing and View. Using these three factors, we can describe a system based on what type of data should be visualized, how it should be visualized, at what point in the surgery it should be visualized and how the user can interact with the data both in terms of manipulation on screen and hardware devices for interaction. Differences between the DVV taxonomy and other visualization taxonomies lie in its specificity to mixed reality visualization in IGS and its ability to account for the constraints of the OR and of the user.

The DVV taxonomy facilitated a complete analysis of 15 state-of-the-art mixed reality IGS systems. The analysis of the publications brought to light particular patterns, for example, the focus on in-situ visualization and the lack of focus in IGS on visualization processing and the interaction component of the view. In doing so, the DVV taxonomy was shown to be useful for finding holes in current research. It therefore, serves both as guide to structure the work that has been done to date in this domain and as a tool for suggesting avenues of future study in the field. The DVV taxonomy was shown to be useful for comparing, analyzing and evaluating systems in a consistent manner and therefore, may help bring us a step toward ensuring the successful introduction of more mixed reality IGS systems in the OR.

The next step of our work will be to evaluate the DVV taxonomy and use it to organize a literature review of mixed-reality IGS systems. By doing so it will become more evident what components of current systems have been shown to improve surgery and can be re-used and which elements need further research and novel solutions.

References

1. Archip, N., Clatz, O., Whalen, S., Kacher, D., Jolesz, F., Golby, A., Black, P.M., Warfield, S.K.: Non-rigid Alignment of Preoperative MRI, fMRI and DT-MRI with Intraoperative MRI to Enhance Visualization and Navigation in Image-guided Neurosurgery. J. Biomech. 39(1), S574 (2006)
2. Bichlmeier, C., Wimmer, F., Heining, S.M., Navab, N.: Contextual Anatomic Mimesis Hybrid In-Situ Visualization Method for Improving Multi-Sensory Depth Perception in Medical Augmented Reality. In: ISMAR, pp. 1–10 (2007)
3. Birkfellner, W., Figl, M., Matula, C., Hummel, J., Hanel, R., Imhof, H., Wanschitz, F., Wagner, A., Watzinger, F., Bergmann, H.: Computer-enhanced Stereoscopic vision in a Head-mounted Operating Binocular. Phys. Med. Biol. 48(3), 49–57 (2003)

4. Blackwell, M., Nikou, C., DiGioia, A., Kanade, T.: An Image Overlay System for Medical Data Visualization. Med. Image. Anal. 4, 67–72 (2000)
5. Das, M., Sauer, F., Schoepf, U.J., Khamene, A., Vogt, S.K., Schaller, S., Kikinis, R., van Sonnenberg, E., Silverman, S.G.: Augmented Reality Visualization for CT-guided Interventions: System Description, Feasibility, and Initial Evaluation in an Abdominal Phantom. Radiology 240(1), 230–235 (2006)
6. Edwards, P.J., King, A.P., Maurer Jr., C.R., de Cunha, D.A., Hawkes, D.J., Hill, D.L., Gaston, R.P., Fenlon, M.R., Jusczyzck, A., Strong, A.J., Chandler, C.L., Gleeson, M.J.: Design and Evaluation of a Aystem for Microscope-assisted Guided Interventions (MAGI). IEEE Trans. Med. Imaging 19(11), 1082–1093 (2000)
7. Gering, D.T., Nabavi, A., Kikinis, R., Hata, N., Donnell, L., Grimson, W., Eric, L., Jolesz, F., Black, P., Wells, W.M.: An Integrated Visualization System for Surgical Planning and Guidance Using Image Fusion and an Open MR. J. Mag. Reson. Imaging 13, 967–975 (2001)
8. Glossop, N., Wang, Z.: Laser Projection Augmented Reality System for Computer-Assisted Surgery. In: CARS 2003, vol. 1256, pp. 65–71 (2003)
9. Hansen, C., Wieferich, J., Ritter, F., Rieder, C., Peitgen, H.-O.: Illustrative Visualization of 3D Planning Models for Augmented Reality in Liver Surgery. Int. J. Comput. Assist. Radiol. Surg. 5(2), 122–141 (2010)
10. Linte, C.A., Moore, J., Wiles, A.D., Wedlake, C., Peters, T.M.: Virtual Reality-enhanced Ultrasound Guidance: A Novel Technique for Intracardiac Interventions. Comput. Aided. Surg. 13(2), 82–94 (2008)
11. Milgram, P., Colquhoun, H.: A Taxonomy of Real and Virtual World Display Integration. In: Mixed Reality - Merging Real and Virtual Worlds, pp. 1–16 (1999)
12. Nicolau, S., Garcia, A., Pennec, X., Soler, L., Ayache, N.: An Augmented Reality System to Guide Radio-frequency Tumour Ablation. Comp. Anim. Virtual World 16(1), 1–10 (2005)
13. Paul, P., Fleig, O., Jannin, P.: Augmented Virtuality Based on Stereoscopic Reconstruction in Multimodal Image-Guided Neurosurgery: Methods and Performance Evaluation. IEEE Trans. Med. Imaging 24(11), 1500–1511 (2005)
14. Sielhorst, T., Bichlmeier, C., Sheining, S.M., Navab, N.: Depth Perception – A Major Issue in Medical AR: Evaluation Study by Twenty Surgeons. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 364–372. Springer, Heidelberg (2006)
15. Stetten, G., Chib, V.: Overlaying ultrasonographic images on direct vision. J. Ultrasound Med. 20(3), 235–240 (2001)
16. Wacker, F.K., Vogt, S., Khamene, A., Jesberger, J.A., Nour, S.G., Elgort, D.R., Sauer, F., Duerk, J.L., Lewin, J.S.: An Augmented Reality System for MR Image-guided Needle Biopsy: Initial Results in a Swine Model. Radiology 238, 497–504 (2006)

Three-Dimensional Ultrasound Probe Pose Estimation from Single-Perspective X-Rays for Image-Guided Interventions

Pencilla Lang[1,2], Petar Seslija[1], Damiaan F. Habets[1,3], Michael W.A. Chu[4,5], David W. Holdsworth[1,3], and Terry M. Peters[1,2,3]

[1] Imaging Research Laboratories, Robarts Research Institute
[2] Biomedical Engineering Graduate Program, University of Western Ontario
[3] Department of Medical Biophysics, University of Western Ontario
[4] Division of Cardiac Surgery, Schulich School of Medicine and Dentistry
[5] London Health Sciences Centre
London, ON, Canada

Abstract. Registration and fusion of real-time trans-esophageal (TEE) and live fluoroscopy images have potential to provide improved visualization during minimally invasive aortic valve replacement and other cardiac interventions. We present an approach for achieving this registration using 3D TEE pose estimation from single-perspective x-ray images, and assess the performance of both point-based and intensity-based tracking techniques. Simulated and experimental accuracy studies demonstrated tracking errors of under 0.58mm and 0.32°, and phantom trials resulted in a 3D US-to-fluoroscopy RMS registration error of 3.5mm.

1 Introduction

Aortic valve replacement is a common procedure that has been performed for decades with good results [1]. Minimally invasive valve implantation techniques have been developed to treat patients previously deemed inoperable. In these procedures, a valve is mounted onto a catheter and guided to position primarily using intra-operative fluoroscopy and transesophageal (TEE) ultrasound. While the fluoroscopy image provides clear visualization of the stent, anatomic structures are visible only on ultrasound. Poor visualization of the coronary ostia in fluoroscopy can lead to the rare but serious complication of coronary obstruction [2]. Fluoroscopy-to-ultrasound registration would enhance guidance by providing a common coordinate frame within which both modalities can be viewed. Initial clinical experience suggests that a 3D localization accuracy of 2 to 3 mm is required.

Previous work has focused on registering the images using magnetic tracking systems [3]. However, magnetic tracking systems are susceptible to field distortions by the presence of metallic materials and electronics, which negatively affects their reliability and tracking accuracy [3] [4].

Since the workflow in the operating room precludes the use of biplane fluoroscopy to localize the probe by triangulation, we propose a fluoroscopy-to-US

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 344–352, 2010.

registration using 2D-to-3D image-to-model registration techniques to determine the 3D pose of the TEE probe from single-perspective fluoroscopy images. Accuracy of an ultrasound-to-fluoroscopy registration depends on the accuracy of both probe tracking in 3D space and ultrasound calibration. In this paper we characterize the accuracy of two fluoroscopy-based tracking techniques when applied to a TEE probe, and present a flouroscopy-US registration based on this technique.

2 2D-to-3D Registration Techniques

Two fluoroscopy-based tracking techniques are employed in this study to localize a TEE probe from a single-perspective x-ray image: a "point-based" and "intensity-based" technique.

Point-based Tracking. Point-based 2D-to-3D registration aligns a rigid-body with radio-opaque markers with corresponding projections within a 2D fluoroscopic image, in a least-squares sense. Details regarding the implementation of this registration are described by Habets *et al* [5], and Hoffman *et al* [6]. This technique requires adding radio-opaque markers to the probe, which can be embedded within the outer plastic casing. Alternatively, an extension at the end of the probe may be tracked. Our clinical experience suggests that an attachment up to 4 cm in length can be safely used without damaging the esophagus during manipulation of the probe.

Intensity-based Tracking. Intensity-based 2D-to-3D registrations align a CT volume with corresponding fluoroscopic images by optimizing an image-similarity metric. Ray-casting is used to project through the CT volume, and create a digitally reconstructed radiograph (DRR). This DRR, or simulated x-ray image, is compared to the actual fluoroscopic image using the gradient difference similarity metric. Details are described by Penny *et al* [7].

3 Simulation Study

Performance of the intensity-based algorithm depends highly on the tracked object's geometry. In this section we describe a method to estimate tracking performance of a given geometry using autocorrelation graphs. This allows optimization of tools for image-based tracking. We apply this technique to the native TEE probe geometry, and to a potential tracking attachment geometry designed to enhance tracking performance.

3.1 Methods

Autocorrelation graphs for each geometry are created by translating the geometry along, or rotating around each axis. The similarity is calculated between

the projected image at each point and a reference projection taken at the reference position (Figure 1a). The peak of this graph represents a correct alignment. As misalignment increases, the similarity metric decreases. The performance of different geometries can be assessed by comparing the shape of the autocorrelation graphs. Narrow-peaked graphs are preferable to wide graphs. The distance between peaks on the autocorrelation graph represents the region in which the tracking algorithm must be initialized for convergence to the correct solution. The shape of the autocorrelation graph varies depending on the noise level.

Ideally, the tracking algorithm will optimize results until the peak of the graph is reached. Actual convergence is determined by the search pattern of the optimization algorithm, the step size, the resolution of the model and images, the geometry of the tracked object, and the presence of noise and calibration errors. The tracking precision in each direction is determined by the position of the model at convergence (Figure 1a).

To determine maximum tracking precision for a given geometry, we register our CT volume to ideal DRR images generated from the same CT. The similarity metric value at convergence is then used to determine tracking precision for each direction from the autocorrelation graph (Figure 1a). The algorithm is initialized to a random position within 2mm and 2 degrees of the actual pose to ensure convergence to the proper peak. We used the Insight-Toolkit (itk) implementation of the gradient descent optimizer, with a step size of 0.001mm [8]. Image and model resolutions are described below.

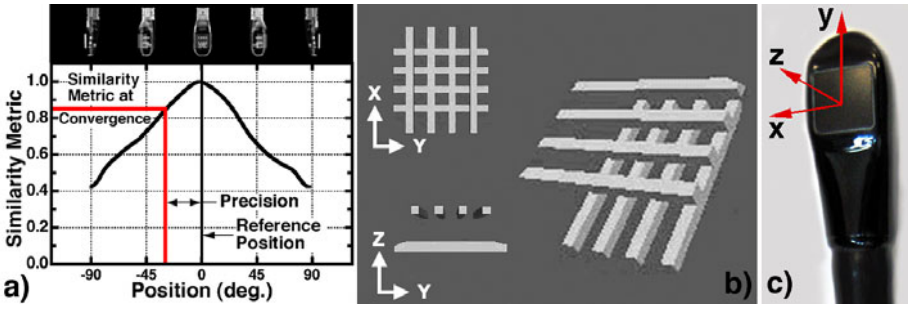


Fig. 1. a) Sample autocorrelation graph demonstrating how tracking error is determined. Actual results are reported in Table 1. b) CAD model of tracking pattern c) TEE Probe with reference coordinate system: In-plane directions = x and y translation, z rotation, out-of-plane directions = z translation, x and y rotations.

Probe Geometry. A micro-CT (eXplore Locus Ultra, GE Healthcare) scan of the US transducer was acquired (0.150mm isotropic voxel spacing, 120kVp, 20mA) from which DRRs were generated. Autocorrelation graphs using a CT of the TEE probe were generated for all 6 degrees of freedom (Figure 1c), for both ideal and noisy images. A Gaussian noise level of 2.5%, representative of clinical images collected on our system, was used.

Tracking Attachment Geometry. Autocorrelation graphs were also generated to assess the performance of a custom-designed tracking attachment pattern. This pattern consists of flat stripes (0.45mm x 8.6mm x 0.151 mm) arranged in two layers offset by 90° (Figure 1b). This pattern is designed to maximize information provided by edges and intensity differences when rotated and translated in the out-of-plane directions. A 3D model was generated using Solidworks^{TM}CAD software (Dassault Systèmes Solidworks Corp). The surface model was converted into a representative CT volume. DRRs created from this volume were used to generate the autocorrelation graphs.

3.2 Results and Discussion

The similarity metric at convergence for both patterns was 0.99 and 0.84 when registering to ideal DRRs, and noisy DRRs, respectively. The tracking errors are reported as displacement error of the probe (mm or degrees), and as the equivalent displacement error of a point on an US plane imaged with the probe (Table 1). For this analysis, we have chosen a representative point: offset 7cm vertically and 3cm horizontally from the image midline. This point approximates common locations of relevant anatomy on clinical images.

In-plane translations and rotations demonstrated high tracking accuracy, with narrow peaks for both ideal and noisy cases, and expected tracking accuracies of less than 0.01mm and 0.1°. Out-of-plane translations and rotations performed well in the ideal case, but were significantly affected by the addition of noise (Table 1). Ideal and 2.5% noise images resulted in maximum US displacement errors of 0.16mm and 6.45mm respectively, significantly higher than clinically acceptable values. This suggests native probe geometry can be used to estimate target location, but not in applications requiring high accuracy.

Tracking attachment geometry performed better than the native probe geometry. Displacement error was reduced nearly tenfold to 0.60mm in noisy images when rotating around the x and y axes, without adversely affecting the other degrees of freedom (Table 1).

Table 1. Expected displacement error in each direction. Translations are given in mm and rotations are given in degrees.

	X Trans	Y Trans	Z Trans	X Rot	Y Rot	Z Rot	Max.US Displacement
Probe – Ideal	0.0025	0.0047	0.0093	0.064	0.063	0.016	0.16
Probe –Noisy	0.49	0.49	2.04	2.64	2.26	0.34	6.45
Pattern –Noisy	0.10	0.12	1.16	0.25	0.71	0.33	0.60

4 Experimental Tracking Study

4.1 Methods

In vitro experiments were performed to assess the accuracy of the point-based and intensity-based tracking techniques. Known displacements (translations or

rotations) were applied to the TEE probe, and compared to the displacement measured from fluoroscopic images. Accuracy was quantified using the root-mean-square (RMS) difference between the measured and applied displacements.

For point-based tracking, a 15mm x 15mm x 20mm rigid-body attachment containing seven spherical tantalum (radius 0.5mm) markers was attached to the end of a Philips X7-2t TEE ultrasound transducer. A micro-CT (eXplore spec-CZT, GE Healthcare) scan of the transducer, with the rigid-body attachment, was acquired (0.050mm isotropic voxel spacing, 110kVp, 32mA) to establish the rigid-body model. Marker locations within the CT volume were measured using region growing and centroiding operations available on micro-CT analysis software (MicroView, GEHealthcare). For intensity-based tracking, the previously acquired microCT was again used (Section 3).

The *in vitro* experiment was performed in a clinical setting using a floor-mounted C-arm radiography system equipped with an X-ray Image Intensifier (Axiom Artis, Siemens Medical). A de-warping grid was employed to correct distortions [9]. Perspective geometry of the radiography system was determined using radiostereometric analysis techniques, consisting of imaging a calibration cage, from which a perspective transform and the focal point are calculated [10].

Using custom-built attachments, the probe was mounted onto both a linear translation table (J. A. Noll Co,^{TM}precision $\pm$ 0.003mm) and a rotational table (Aerotech Inc^{TM}Model MR100, precision $\pm$ 5E-4°), which provided gold standard displacement measurements (Figure 2a). Each of the 6 degrees of freedom (x, y, z translation and rotation) were assessed independently. 16-24 displacements were applied in each direction, with steps sizes ranging from 0.635mm to 5.08mm and 1 to 10°. A fluoroscopy image (isotropic pixel spacing 0.19mm, 58 kVp, 23 mA, 4ms) was acquired both prior and subsequent to each displacement(Figure 2b). Both the native geometry of the US probe and the tracking attachment are visible in all the fluoroscopy images, allowing the same images to be used for both the point-based and intensity-based accuracy assessment (the point-based markers were excluded from the intensity-based ROI). The acquired images were processed offline on a personal computer with a 2.6GHz

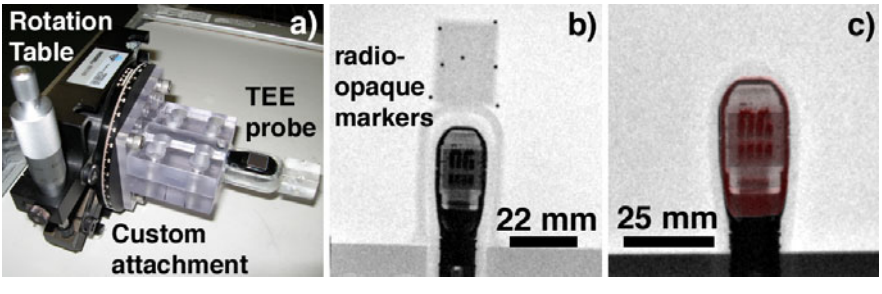


Fig. 2. a) Experimental setup b) Fluoroscopy image showing probe and markers c) Co-registered DRR (red) and fluoroscopic image from intensity-based tracking

quad core CPU and 4GB of memory. The point-based algorithm requires manual initialization to establish correspondence between the rigid-body points and their projections within the fluoroscopy image. The intensity-based algorithm is initialized by defining a region of interest (excluding the radio-opaque markers), and manually estimating the initial registration parameters to provide the algorithm with a starting point. Since the position of the probe is constrained within the esophagus, it will be relatively easy to automate this initialization process.

4.2 Results and Discussion

Maximum tracking errors associated with the point-based and intensity-based tracking were 0.58mm, 0.32° and 2.29mm, 3.76°, respectively (Table 2, Figure 2c). Both techniques demonstrated a higher accuracy tracking in-plane movements compared to out-of-plane movements. This decreased accuracy is explained by the limited sensitivity of information available out-of-plane (the geometric magnification). The point-based technique demonstrated greater accuracy than a 2D TEE probe mounted with an external magnetic sensor as described by Moore *et al* [4]. The results of the intensity-based tracking are consistent with simulation results presented in Section 1, where tracking using only the native probe geometry demonstrated limited accuracy. This suggests that accurate fluoroscopy-US registration requires the use of a custom geometry tracking attachment. Although our results from Section 1 suggest that optimized geometries may improve performance of the intensity-based registration, we found that the point-based was more accurate, easier to implement, and has significantly faster registration times.

Table 2. Root mean square error in each direction. Translations are given in mm and rotations are given in degrees.

	X Trans	Y Trans	Z Trans	X Rot	Y Rot	Z Rot
Point-based	0.21	0.21	0.58	0.32	0.13	0.16
Intensity-based (native probe geometry)	0.52	0.30	2.29	3.47	3.76	1.05

5 Fluoroscopy-US Registration

5.1 Methods

Fluoroscopy and US images (iE33, Philips Healthcare) were collected simultaneously, using the US probe and C-arm systems described in Section 3. The US probe was visible in all acquired fluoroscopy images; allowing the location of the probe in world space to be determined using the point-based registration technique described in Section 2. The 3D world location of points within the 2D

US image are given by Equation 1, where the points and transformations are defined as follows:

P_{US}, P_{World} = Point locations in 2D US image space and 3D world space
$T_{US \rightarrow Probe}$ = Transformation from 2D US image space to probe space
$T_{Probe \rightarrow World}$ = Transformation from probe space to 3D world space

$$P_{world} = T_{Probe \rightarrow World} \bullet T_{US \rightarrow Probe} \bullet P_{US} \quad (1)$$

US probe calibration was performed using the Z-bar technique described by Gobbi *et al* [11] to determine $T_{US \rightarrow Probe}$. Radio-opaque markers were attached to the Z-bar phantom, and their relative locations characterized, enabling it to be registered to world space; this allows all measurements to be made in a common frame of reference. The phantom was immersed in 10.5% glycerol at 21° Celsius.

Fluoroscopy-US registration error was assessed using a target phantom consisting of a table-tennis ball mounted on a frame (Figure 3a). Radio-opaque markers were added to the frame, allowing the "ground truth" cross section to be defined using microCT data (eXplore Locus Ultra), and point-based pose estimation (Section 2). Six cross-sectional US images of the ball were acquired at varying angles and locations used in clinical practice. The centroid of the cross-section was identified using a least squares best fit ellipse to manually identified points (Figure 3b). Both 3D (world space) and 2D (fluoroscopy image) target registration errors were calculated.

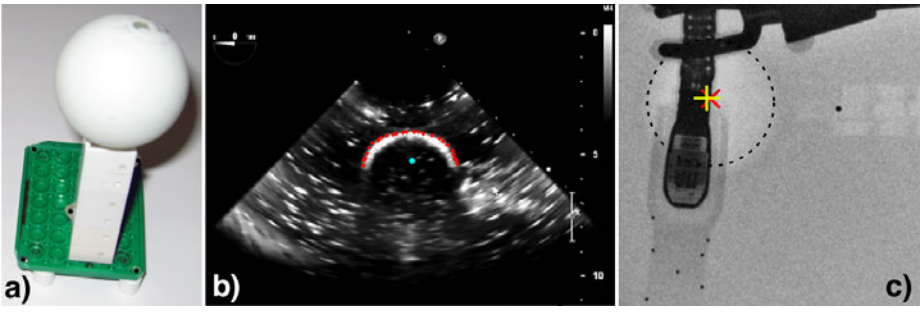


Fig. 3. a) Target phantom b) Centroided US Image c) Gold standard (red), Measured (yellow) targets

5.2 Results and Discussion

Z-bar calibration was achieved with a RMS error of 1.6 mm, consistent with calibration errors reported in similar studies [3]. An RMS target registration error of 3.5mm was achieved when localizing US points in 3D world space, which reduces to 2.9mm when reprojected onto the fluoroscopy image (Table 3, Figure 3c). These results are similar to those obtained with magnetic tracking [3] [4], and consistent with expected results based on the inherent probe tracking error

Table 3. Target registration errors, given in mm.

	RMS	Mean	Std Dev
World Space	3.47	3.42	0.62
Fluoroscopy Image	2.91	2.77	0.98

described in Section 4. Target registration error could be significantly reduced with a more accurate and robust US probe calibration. Alternative calibration techniques can be adapted from magnetic to fluoroscopy tracking [3] [12].

6 Conclusions

Registration and fusion of real-time TEE and live fluoroscopy images have the potential to provide improved visualization during minimally invasive off-pump aortic valve replacement and other interventional cardiac procedures. We present an approach for achieving this registration using 3D TEE pose estimation from single-perspective x-ray images. This tracking technique provides improved accuracy in probe localization over magnetic tracking, and does not significantly change the workflow of the operating room. Experimental assessment of tracking accuracy demonstrated maximum RMS displacement and rotation errors of 0.58mm, 0.32° and 2.29mm, 3.76° for point-based and intensity-based tracking techniques, respectively. These results are consistent with our simulation study, which estimated a maximum tracking accuracy of 2.04mm and 2.64° for ideal images with similar noise levels. The point-based method is most appropriate for situations requiring high accuracy, while the intensity-based method provides the advantage of using native probe geometry. Combined with Z-bar US calibration, a 3D RMS registration accuracy of 3.5mm was achieved. Future work will focus on improving the US calibration procedure, and evaluating clinical feasibility through *in vivo* animal trials.

References

1. Walther, T., Simon, P., et al.: Transapical minimally invasive aortic valve implantation: multicenter experience. Circulation 116(11 Suppl.), I240–I245 (2007)
2. Lichtenstein, S.V., Cheung, A., et al.: Transapical transcatheter aortic valve implantation in humans: initial clinical experience. Circulation 114(6), 591–596 (2006)
3. Jain, A., Gutierrez, L., Stanton, D.: 3D TEE Registration with X-Ray Fluoroscopy for Interventional Cardiac Applications. In: Ayache, N., Delingette, H., Sermesant, M. (eds.) FIMH 2009. LNCS, vol. 5528, pp. 321–329. Springer, Heidelberg (2009)
4. Moore, J.T., Wiles, A.D., Wedlake, C., et al.: Integration of trans-esophageal echocardiography with magnetic tracking technology for cardiac interventions. In: SPIE, vol. 7625, 7625Y (2010)
5. Habets, D.F., Holdsworth, D.W., et al.: Error analysis of marker-based object localization using a single-plane XRII. Med. Phys. 36(1), 190–200 (2009)
6. Hoffmann, K.R., Esthappan, J.: Determination of three-dimensional positions of known sparse objects from a single projection. Med. Phys. 24(4), 555–564 (1997)

7. Penney, G.P., et al.: A comparison of similarity measures for use in 2-D-3-D medical image registration. IEEE Trans. Med. Imaging 17(4), 586–595 (1998)
8. Yoo, T., Whitaker, R., et al.: Engineering and Algorithm Design for an Image Processing API: A Technical Report on ITK - The Insight Toolkit. In: Proc. of Medicine Meets Virtual Reality (2002)
9. Fahrig, R., Holdsworth, D.W., et al.: Three-dimensional computed tomographic reconstruction using a C-arm mounted XRII: correction of image intensifier distortion. Med. Phys. 24(7), 1097–1106 (1997)
10. Selvik, G., et al.: A roentgen stereophotogrammetric system. Construction, calibration and technical accuracy. Acta Radiol. Diagn (Stockh) 24(4), 343–352 (1983)
11. Gobbi, D.G., Comeau, R.M., Peters, T.M.: Utrasound Probe Tracking for Real-Time Ultrasound/MRI Overlay and Visualization of Brain Shift. In: Taylor, C., Colchester, A. (eds.) MICCAI 1999. LNCS, vol. 1679, pp. 920–927. Springer, Heidelberg (1999)
12. Chen, T., Abolmaesumi, P., et al.: A real-time ultrasound calibration system with automatic accuracy control and incorporation of ultrasound section thickness. In: Proc. SPIE, vol. 6918, 6918A (2008)

Automated Nomenclature of Upper Abdominal Arteries for Displaying Anatomical Names on Virtual Laparoscopic Images

Kensaku Mori[1,2,⋆], Masahiro Oda[1], Tomohiko Egusa[2], Zhengang Jiang[2], Takayuki Kitasaka[3], Michitaka Fujiwara[4], and Kazunari Misawa[5]

[1] Information and Communications Headquarters, Nagoya University, Japan
kensaku@is.nagoya-u.ac.jp, moda@mori.m.is.nagoya-u.ac.jp
[2] Graduate School of Information Science, Nagoya University, Japan
[3] Faculty of Information Science, Aichi Institute of Technology, Japan
[4] Graduate School of Medicine, Nagoya University, Japan
[5] Aichi Cancer Center Hospital, Japan

Abstract. This paper presents a method for automated nomenclature of abdominal arteries that are extracted from 3D CT images based on the combination optimization approach for the displaying anatomical names on virtual laparoscopic images. It is important to understand the blood vessel network of a patient. Our proposed method recognizes the anatomical names of each arterial branch extracted from contrasted 3D images based on geometric features. We employ a combination optimization approach for treating the variations of branching patterns and overlay recognized anatomical names on virtual laparoscopic views for assisting the recognition of patient anatomy for surgeons. Experimental results using 89 cases of 3D CT images showed that the nomenclature accuracy for uncorrected blood vessel tree and corrected blood vessel tree were about 84.2% and 88.8% in average respectively and demonstrated anatomical name overlay on virtual laparoscopic images.

1 Introduction

In laparoscopic surgery, surgeons must understand the patient anatomy, especially the blood vessels. Abdominal vasculature is quite complex, and there are several variations, such as branching orders, locations or the existence of missing or additional branches. Such complexities in abdominal vasculature may increase surgeons' loads or affect the safety of the surgery.

To overcome such problems, CT images are taken for understanding the patient anatomy prior to laparoscopic surgery. Surgeons make surgical plans by

⋆ The authors thank our colleagues for suggestions and advices. Parts of this research were supported by the Grant-In-Aid for Scientific Research from the Ministry of Education (MEXT), Japan Society for the Promotion of Science (JSPS), the Japan Society for Promotion of Science, and the Grant-In-Aid for Cancer Research from the Ministry of Health and Welfare.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 353–362, 2010.

reading or visualizing these CT images. However, since the abdominal vasculature has a complex network, it is hard to understand the vasculatures by CT images. Assistance systems for grasping abdominal blood vessels are expected to be developed.

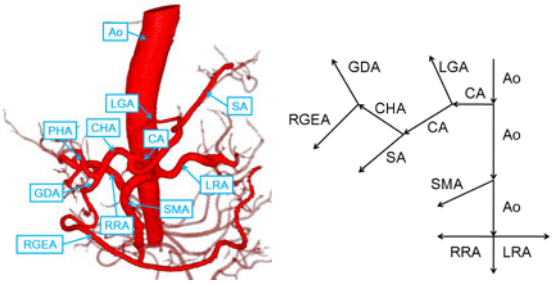
This paper presents a method for automated nomenclature (automated anatomical labeling) of abdominal arteries that are extracted from 3D CT images based on a combination optimization approach for displaying anatomical names on virtual laparoscopic (VL) images. Automated nomenclature is partially implemented using a machine learning approach.

There are several researches on automated nomenclature. Mori et al. presented a method for automated nomenclature of bronchial branches extracted from CT images [1,2]. Tschirren et al. also presented such a method for human airway tree based on the branchpoint matching algorithm [3]. The most difference between bronchi and artery nomenclature is that the abdominal artery nomenclature process needs to consider that arteries are branching from thick arteries. On the other hand, the bronchus shows equal branching at every branching point (one branch bifurcates into two branches of the same size.) Since the existing automated nomenclature algorithms for the bronchus do not work well for artery nomenclature, special procedures for automated artery nomenclature need to be developed. Hence, direct comparison with other methods including Mori's latest is difficult. Chalopin et al. introduced 3D models of coronary arteries for automated nomenclature and tried nomenclature of coronary arteries on 2D X-ray images [4]. Won et al. presented an interesting method for the uncluttered single-image generation of abdominal arteries[5]. Although they analyzed the branching structures of abdominal arteries, automated nomenclature was not performed. To our knowledge, there is no research on the anatomical name display of abdominal arteries on VL views based on automated nomenclature.

2 Target Arteries and Problem Formulation

(a) Target arteries. In this paper, we selected eleven upper abdominal arteries as the target arteries of automated nomenclature and the anatomical name display on the VL views (Fig. 1.) Upper abdominal arteries are crucial for surgery of the stomach, the liver, the pancreas, and the gallbladder. The target arteries are: (a) the abdominal aorta (**Ao**), (b) the celiac artery (**CA**), (c) the common hepatic artery (**CHA**), (d) the splenic artery (**SA**), (e) the proper hepatic artery (**PHA**), (f) the gastroduodenal artery (**GDA**), (g) the right gastroepiploic artery (**RGEA**), (h) the left gastric artery (**LGA**), (i) the superior mesenteric artery (**SMA**), (j) the left renal artery (**LRA**), and (k) the right renal artery (**RRA**). Abbreviations of these words, which are commonly used in the medical field, are shown in the brackets.

(b) Problem formulation. In the proposed method, the abdominal artery network is represented by a graph. The branches and the branching points of the blood vessels are represented as edges and nodes, respectively. Here, we consider



(a) (b)

Fig. 1. Illustration of abdominal arteries for automated nomenclature: (a) 3D views of abdominal arteries and (b) branching structure. See Section 2 for abbreviations of anatomical names.

the automated nomenclature process an optimization problem that selects the most appropriate combination of pairs of branches and anatomical names. This process can be expressed as

$$\mathbf{c}_* = \underset{\mathbf{c}_k \in \mathbf{S}}{\operatorname{argmax}}\ h(\mathbf{c}_k), \tag{1}$$

where $\mathbf{c}_k$ means a combination list that enumerates all possible pairs of blood vessel branches and their associated anatomical names. $\mathbf{S}$ is a set of all possible combination lists. $h(\mathbf{c}_k)$ is a function that evaluates $\mathbf{c}_k$. k-th combination list $\mathbf{c}_k$ can be expressed as

$$\mathbf{c}_k = \{(\mathbf{b}_i,\ j) \mid \mathbf{b}_i \in \mathbf{B},\ j \in \mathbf{L}\}, \tag{2}$$

where $\mathbf{B}$ shows a set of blood vessel branches (edges in graph structure; concatenated branches are also considered) and $\mathbf{L}$ represents a set of anatomical names. For example, $\mathbf{c}_k$ becomes $\mathbf{c}_1 = \{(b_1, \mathbf{Ao}), (b_2, \mathbf{CA}), (b_3, \mathbf{SA}), \cdots\}$.

To maximize the automated nomenclature performance and to reduce the computation time, we divide the abdominal arteries into four parts and sequentially perform specific nomenclature procedures (specific evaluation functions) for each part. The four parts consist of: (a) {**Ao**}, (b) {**CA, CHA, SA, PHA, GDA+RGEA, LGA**}, (c) {**SMA**}, and (d) {**LRA, RRA**}. The nomenclature process is executed in this group order. In other words, first we assign **Ao** and then assign {**CA, CHA, SA, PHA, GDA+RGEA, LGA**}. The nomenclature of {**SMA**} and {**LRA, RRA**} is then executed. A minimization process is executed for each group. Anatomical (i.e. parent-and-child relationship) and shape constraints are also considered for reducing the number of combinations.

(c) Features of abdominal artery branch. The proposed method assigns anatomical names based on the features calculated for each vessel branch. Blood vessel branch $\mathbf{b}_i$ possesses the following features: (a) running direction $\mathbf{d}_i$, (b) norm n_i of $\mathbf{d}_i$, (c) length l_i, and (d) diameter $\mathbf{t}_i$. Running direction $\mathbf{d}_i$ can be computed as $\mathbf{d}_i = \mathbf{p}_i^{\mathbf{e}} - \mathbf{p}_i^{\mathbf{s}}$, where $\mathbf{p}_i^{\mathbf{s}}$ and $\mathbf{p}_i^{\mathbf{e}}$ mean the start and end points

of branch $\mathbf{b}_i$, respectively. Norm n_i is calculated as $n_i = \|\mathbf{d}_i\|$. Length l_i can be computed along branch $\mathbf{b}_i$ by a 6-neighborhood distance measure along the branch. Diameter $\mathbf{t}_i$ can be obtained by referring to the result of the Euclidean distance transformation of the abdominal artery region.

3 Methods

3.1 Overview

The basic idea of the proposed method is to assign anatomical names to each branch based on features defined in the previous section. We enumerate the possible combination list of holding pairs of branches and the associated anatomical names and select the best combination list based on the features.

3.2 Preprocessing

We extracted the artery regions from the contrasted CT images by utilizing a method presented by Nakamura et al. [7] that basically enhances the blood vessels by utilizing the eigenvalues of a Hessian matrix [6] and then extracts the blood vessel regions by thresholding followed by connected component analysis. The thinning algorithm was applied to the extracted regions to obtain the centerlines of the abdominal arteries. A graph representation is obtained by the method shown in [1].

3.3 Automated Nomenclature

(a) Labeling of Ao: Since the abdominal aorta is the thickest blood vessel inside the human body, it is possible to find it by considering its diameter. For abdominal aorta nomenclature, we define function h as the function that finds combination $\mathbf{c}_k$, where (a) the diameters of all branches $\mathbf{b}_i$ are bigger than threshold T_{Ao} mm, (b) all candidate labels are **Ao**, and (c) all branches are connected.

(b) Nomenclature of CA, CHA, SA, PHA, GDA+RGEA, and LGA: For this part, we utilize a machine learning approach for nomenclature. In the learning step, we compute feature vectors $\mathbf{x}_i = \{\widehat{d}_i^x, \widehat{d}_i^y, \widehat{d}_i^z, n_i, l_i\}$ and manually assign anatomical names for all branches $\mathbf{b}_i$, where $\widehat{d}_i^x$, $\widehat{d}_i^y$, and $\widehat{d}_i^z$ are the x, y, and z components of normalized vector $\widehat{\mathbf{d}}_i$ of running direction $\mathbf{d}_i$ of branch $\mathbf{b}_i$, and n_i norm of $\mathbf{d}_i$, respectively. For each anatomical label a_j, we compute mean $\mathbf{m}_j$ and covariance matrix $\mathbf{\Sigma}_j$ from features vector $\mathbf{x}_i$ of the branch whose anatomical name is j.

In the testing step, we compute an evaluation value for combination list $\mathbf{c}_k$ by

$$h(\mathbf{c}_k) = \sum_{i}^{N} P_{i,j(i)}, \tag{3}$$

where $j(i)$ is the candidate label for branch i in $\mathbf{c}_k$ and N is the number of branches in the combination list. $P_{i,j(i)}$ is the likelihood that shows that branch $\mathbf{b}_i$ has anatomical name j and is computed by

$$P_{i,j(i)} = p_{j(i)}(\mathbf{x}_i) \tag{4}$$

$$= \frac{1}{(2\pi)^3|\mathbf{\Sigma}_{j(i)}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}_j - \mathbf{m}_{j(i)})^T(\mathbf{\Sigma}_{j(i)})^{-1}(\mathbf{x}_j - \mathbf{m}_{j(i)})\right], \tag{5}$$

where $\mathbf{x}_i$ is a feature vector computed for $\mathbf{b}_i$. Selected combination list $\mathbf{c}_*$ using Eq. (1) becomes nomenclature result.

To achieve fast computation and reduce misnomenclature, we introduce the limitation of the combination lists of the branches and the candidate labels in two ways: (a) anatomical constraints and (b) feature constraints. In the first constraint, we consider the parent-and-child relationship in the enumeration of the combination lists. For example, there is an anatomical feature in which **PHA** only branches off from **CHA**. From set **S**, we exclude the combination lists that do not satisfy the anatomical constraints. The second constraint excludes combination lists that have low possibilities to be selected as an optimal combination by checking the features of the branches contained in $\mathbf{c}_k$. If $\mathbf{b}_i$, which will be assigned to anatomical name j, fails to satisfy one of the following conditions

$$\left|\frac{l_i - \mu_j^l}{\sigma_j^l}\right| < \alpha, \ \left|\frac{n_i - \mu_j^n}{\sigma_j^n}\right| < \beta,$$

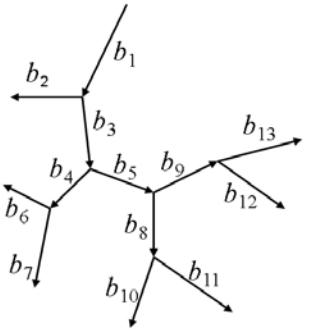
$$\sqrt{(\widehat{\mathbf{d}}_i - \mathbf{m}_j^{\widehat{\mathbf{d}}})^t(\mathbf{\Sigma}_j^{\widehat{\mathbf{d}}})^{-1}(\widehat{\mathbf{d}}_i - \mathbf{m}_j^{\widehat{\mathbf{d}}})} < \gamma, \ \left|\frac{t_i - \mu_j^t}{\sigma_j^t}\right| < \delta, \tag{6}$$

then we exclude $\mathbf{c}_k$, which includes $\mathbf{b}_i$, from **S**. Here, μ_j^l, μ_j^n, $\mathbf{m}_j^{\widehat{\mathbf{d}}}$, and μ_j^t show the means or the mean vector of l_j, n_j, $\widehat{\mathbf{d}}_j$, and t_j of branches with anatomical name j. σ_j^l, σ_j^n, $\mathbf{\Sigma}_j^{\widehat{\mathbf{d}}}$, and σ_j^t are the standard deviations or the covariance matrix. α, β, and γ are the parameters for branch selection.

(c) Nomenclature of SMA: **SMA** branches off from **Ao** and runs to the foot side. Its starting branch can be easily identified as the branch that branches off from **Ao** and runs to the anterior direction. **SMA** has an anatomical feature where its diameter does not change so much, although many thinner branches are branching from it. We define evaluation function h as

$$h(\mathbf{c}_k) = \sum_i^N \frac{1}{\mid t_i - t_{start} \mid} \tag{7}$$

where t_{start} and t_i show the thickness of the starting branch of **SMA** and branch $\mathbf{b}_i$. N is the number of branches in $\mathbf{c}_k$. In the above evaluation, we only consider combinations with candidate label **SMA** and child branches of its starting branch (for example, $\mathbf{c}_k = \{(b_1, \mathbf{SMA}), (b_3, \mathbf{SMA}), (b_4, \mathbf{SMA}), \cdots\}$ and $\mathbf{b}_i$ are the child branches of the starting branch of **SMA**) in many candidate lists. We select the combination list having the highest evaluation value.



(a)

	CA		SA		CHA		PHA		GDA+RGEA		LGA	
	C.B.	L.	C.B.	L.	C.B.	K.	C.B.	L.	C.B.	L.	C.B.	L.
$\mathbf{c_1}$	b_{1+3}	$P_{1+3,1}$	b_4	$P_{4,2}$	b_5	$P_{5,3}$	b_8	$P_{8,4}$	b_9	$P_{9,5}$	b_2	$P_{2,6}$
$\mathbf{c_2}$	b_{1+3}	$P_{1+3,1}$	b_4	$P_{4,2}$	b_5	$P_{5,3}$	b_9	$P_{9,4}$	b_8	$P_{8,5}$	b_2	$P_{2,6}$
$\mathbf{c_3}$	b_{1+3}	$P_{1+3,1}$	b_4	$P_{4,2}$	b_5	$P_{5,3}$	b_9	$P_{9,4}$	b_{8+10}	$P_{8+10,5}$	b_2	$P_{2,6}$
$\mathbf{c_4}$	b_{1+3+5}	$P_{1+3+5,1}$	b_8	$P_{8,2}$	b_9	$P_{9,3}$	b_{12}	$P_{12,4}$	b_{13}	$P_{13,5}$	b_2	$P_{2,6}$
$\mathbf{c_5}$	b_{1+3+5}	$P_{1+3+5,1}$	b_8	$P_{8,2}$	b_9	$P_{9,3}$	b_{12}	$P_{12,4}$	b_{13}	$P_{13,5}$	b_4	$P_{4,6}$
$\mathbf{c_6}$	b_{1+3+5}	$P_{1+3+5,1}$	b_8	$P_{8,2}$	b_9	$P_{9,3}$	b_{12}	$P_{12,4}$	b_{13}	$P_{13,5}$	b_{4+7}	$P_{4+7,6}$
...	...	...	...	...	...	...	...	...	...	...	...	...

(b)

Fig. 2. Example of input graph structure and associated combination lists. (a) graph structure of abdominal artery and (b) combination lists for (a). "C.B." and "L." mean "candidate branch" and "likelihood", respectively. Representation such as "1+3" means concatenated branches. For example, $\mathbf{c}_1$ shows a combination list that assigned CA to b_{1+2}, **SA** to b_4, **CHA** to b_5, and so on.

(d) Nomenclature of LRA and RRA: **LRA** and **RRA** branch off from **Ao**. Their roots are very close to the branching point of **Ao** and **CA**. Hence, we make set of combination lists **S** by selecting the combination lists that include the branches that branch off from the **Ao** obtained in Sec. 3.3(a). We define the condition to determine whether branch $\mathbf{b}_i$ branches off at a point close to **CA** (obtained in Sec. 3.3(b)) or not as

$$\left| \frac{t_i - \mu^t}{\sigma^t} \right| < \epsilon, \tag{8}$$

where $t_i \mathrm{C} \mu^t \mathrm{Cand}\ \sigma^t$ are the Euclidean distances between the starting points of branch $\mathbf{b}_i$ and **CA**, the mean and the standard deviation of t_i computed from the learning datasets. ϵ is a parameter that controls the assignment. Since there are individual variations on the number of **LRA**s and **RRA**s, we define evaluation function h as

$$h(\mathbf{c}_k) = \begin{cases} 1 & \text{if all conditions of Eqs. (6) and (8) are satisfied} \\ 0 & \textit{otherwise}. \end{cases} \tag{9}$$

We select all the combination lists that satisfy $h(\mathbf{c}_k) = 1$. If the branches run to the left direction, they are assigned as **LRA**. If they run to the right direction, they are assigned as **RRA**.

3.4 Anatomical Name Display on VL Images

By using the results of automated nomenclature, we overlay the anatomical names of the abdominal arteries on the VL views. Since the centerlines and the branching structures are computed in the automated nomenclature process, we can find the location of each artery on a VL view by projecting their centers of locations on a projection screen. We display the anatomical names at the projected locations on the VL views and change the sizes and the colors based on to the distance from the viewpoint.

4 Experiments

We applied our proposed method to 89 cases of contrasted abdominal CT images. Image acquisition parameters were 512x512 pixels in the slice planes, 301 to 1061 slices, 0.430 to 0.782 mm pixel spacing, 0.5 mm to 1.0 mm X-ray collimation, and 0.4 to 0.5 mm reconstruction pitch. Although the CT images were taken for surgical planning for gastric cancer resection, no patients had any abnormalities (such as no stenosis) in their abdominal artery structure. We conducted the experiments by the leave-one-out scheme and utilized the following parameters: $T_{Ao} = 5.0$, $\alpha = 4.5$, $\beta = 10.0$, $\gamma = 4.5$, and $\delta = 6.0$ for Section 2.3 (b), and $\alpha = 3.0$, $\beta = 6.0$, $\gamma = 3.0$, and $\epsilon = 3.0$ for Section 2.3 (d). The averages of the automated nomenclature accuracy for all anatomical names were computed. The artery regions extracted from the CT images contain false negative branches, false positive branches or holes caused by noise. For only evaluation purposes, we manually corrected the artery regions and measured the nomenclature accuracy. The ground truth data of the anatomical labels was created under the support of a medical doctor who is also a co-author of this paper. The results of nomenclature are shown in Table 1 and Fig. 3. The computation time for nomenclature for one case takes 10 seconds to 10 minutes; depends on the number of branches. We utilized the automated nomenclature results for the augmentation of the VL views. The results are shown in Fig. 4 and in the supplementary videos. In these figures or videos, we generated VL views by volume rendering and overlaid the anatomical names on the VL images.

5 Discussion

This paper presented a method for nomenclature of abdominal arteries based on features and combination optimization. As shown in Table 1, we performed automated nomenclature with 84.2% accuracy for the non-corrected data and 88.8% for the corrected data by using only features computed on each arterial

Table 1. Accuracy of automated nomenclature for artery regions that were manually corrected and not corrected. Manual correction was performed to delete spurious branches and holes caused by the thinning algorithm

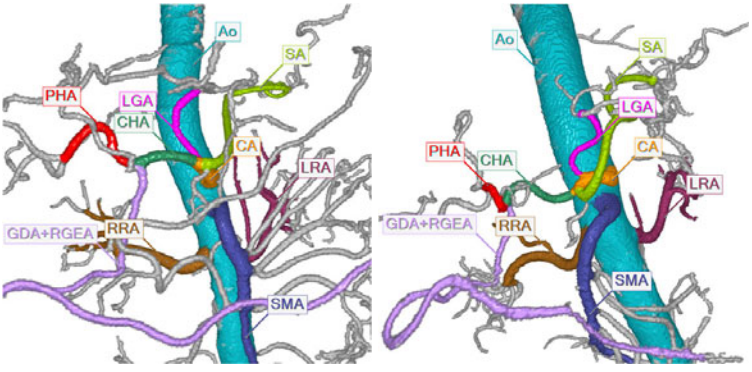
Arteries	Accuracy	
	w/o manual correction	w/ manual correction
Ao	92.1%	100%
CA	87.6%	94.4%
CHA	84.2%	91.0%
SA	88.8%	93.3%
PHA	84.3%	91.0%
GDA + RGEA	84.1%	89.8%
LGA	75.0%	83.0%
SMA	82.0%	87.6%
LRA	74.5%	77.5%
RRA	79.8%	80.9%
Average	84.2%	88.8%

branch. In Fig. 3, several branches, which are considered as branching variations, are branching off from **CA** and **SA**. However, the proposed method was able to assign correct anatomical names to **CA** and **SA**. These are good examples showing advantages of the proposed method utilizing the combination optimization approach.

On the other hand, nomenclature accuracy of **LRA** and **RRA** is obviously lower than other arteries. This could be improved by finding other features that can distinguish the renal and other arteries. The combination lists were eliminated by using anatomical constrains for fast processing. This is because the number of combination lists becomes huge. Although anatomical constrain can greatly reduce the number of combination lists, the proposed method failed to label the branches having minor branching variations. Such misnomenclature can be prevented by adding such branching patterns to anatomical constraints, and good features must also be found that can output high likelihood for such minor branching cases.

This paper extracted artery regions by utilizing the Hessian-based method [6]. However, it caused some holes in artery regions or rough surfaces that generate many spurious branches in the thinned results. Although the nomenclature accuracy for the uncorrected data becomes lower than that of the corrected data, the nomenclature accuracy still remained at 84.2% The combination optimization approach prevented worse results.

As we mentioned in the introduction, there are several researches on automated nomenclature [2,3]. These algorithms were developed only for bronchial branch nomenclature. The largest difference between bronchi and artery nomenclature is that the abdominal artery nomenclature process needs to consider the situation where many arteries are branching from thick (trunk) arteries. These thick branches still retain same anatomical names after branching. On the other hand, the bronchus shows equal branching at every branching point



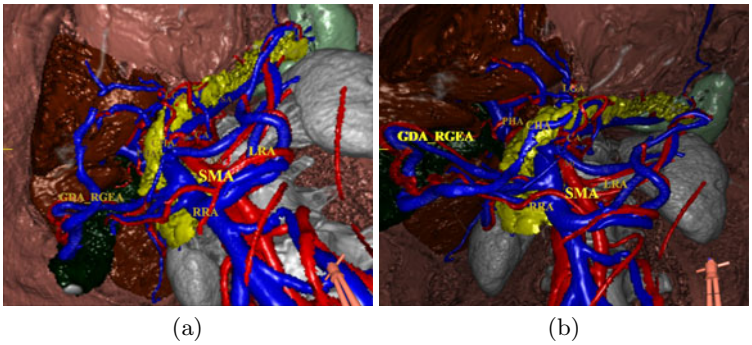
(a) (b)

Fig. 3. Examples of automated nomenclature results

(one branch bifurcates into two branches of the same size.). The anatomical name changes at every branching point. The existing nomenclature algorithms for the bronchus cannot be directly utilized for artery nomenclature. Although the proposed method employs the machine learning approach for nomenclature, as does Mori's method [2], the whole procedure is different. A new algorithm that consider the features of abdominal arteries should be developed. This is the novel part of this paper. This specialty complicates direct comparisons of the proposed method with the existing methods developed for the bronchus, including Mori's latest method [2].

Application of the proposed method to other organs is also challenging. The nomenclature process based on machine learning and combination optimization can be applied to nomenclature of liver vasculature that resembles to abdominal artery branching, if appropriate features and evaluation functions can be defined.

We overlaid anatomical names on VL images. Such anatomical overlay greatly assists surgeons to understand patient anatomy. If we can synchronize VL view



(a) (b)

Fig. 4. Anatomical name overlay on VL views. (a) and (b) are captured at different viewpoints

with real ones and overlay anatomical names on real laparoscopic views, it would greatly assist laparoscopic surgery.

6 Conclusions

This paper presented a method for automated nomenclature of abdominal arteries that were extracted from 3D CT images based on the combination optimization approach for displaying anatomical names on VL images. The experimental results demonstrated nomenclature with accuracy of about 89% and automated display of anatomical names on VL views. Future work includes an extension of the target arteries of nomenclature such as lower abdominal arteries, improvement of the features, and the introduction of a sophisticated machine learning approach such as AdaBoost.

References

1. Mori, K., Hasegawa, J., Toriwaki, J., et al.: Automated anatomical labeling of the bronchial branch and its application to the virtual bronchoscopy system. IEEE Trans. on Medical Imaging 19(2), 103–114 (2000)
2. Mori, K., Ota, S., Deguchi, D., et al.: Automated Anatomical Labeling of Bronchial Branches Extracted from CT Datasets Based on Machine Learning and Combination Optimization and Its Application to Bronchoscope Guidance. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 707–714. Springer, Heidelberg (2009)
3. Tschirren, J., McLennan, G., Palagyi, K., et al.: Matching and anatomical labeling of human airway tree. IEEE Trans. on Medical Imaging 24(12), 1540–1547 (2005)
4. Chalopin, C., Finet, G., Magnin, I.E.: Modeling the 3D coronary tree for labeling purposes. Medical Image Analysis 5, 301–315 (2001)
5. Won, J.H., Rosenberg, J., Rubin, G.D., et al.: Uncluttered Single-Image Visualization of the Abdominal Aortic Vessel Tree: Method and Evaluation. Medical Physics 36(11), 5245–5260 (2009)
6. Sato, Y., Nakajima, S., Shiraga, N., et al.: Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. Medical Image Analysis 2(2), 143–168 (1998)
7. Nakamura, Y., Tsujimura, Y., Kitasaka, T., et al.: A study on blood vessel segmentation and lymph node detection from 3D abdominal X-ray CT images. International Journal of Computer Assisted Radiology and Surgery 1(Suppl. 1), 381–382 (2006)

Hidden Markov Model for Quantifying Clinician Expertise in Flexible Instrument Manipulation

Jagadeesan Jayender[1,2], Raúl San Jośe Estépar[1], Keith Obstein[3], Vaibhav Patil[1,2], Christopher C. Thompson[3,⋆], and Kirby G.Vosburgh[1,2]

[1] Department of Radiology, Harvard Medical School, Brigham and Women's Hospital, Boston, MA, USA
jayender@bwh.harvard.edu, rjosest@bwh.harvard.edu, patil@bwh.harvard.edu, kirby@bwh.harvard.edu

[2] CIMIT Image Guidance Laboratory, Boston, MA,USA

[3] Division of Gastroenterology, Brigham and Women's Hospital, Boston, Massachusetts, USA
kobstein@partners.org, ccthompson@partners.org

Abstract. Clinicians are trained to manipulate a colonoscope while minimizing the force exerted on the colon walls to reduce the danger of luminal perforation and discomfort to the patient. Here, we classify the expertise of the clinician performing colonoscopy using a Hidden Markov Model. Seven models are trained corresponding to the performance of the expert in the entire colon, ascending, transverse and descending colon and three gestures corresponding to roll and two angulations of the distal end of the scope. Experimental results in a colon model (CM-1, Olympus, Tokyo, Japan) are shown to compare the performance of the four groups of users - first year, second year and third year GI residents and expert physicians.

1 Introduction

The performance of gastroenterologists, surgeons, and related practitioners has historically been assessed subjectively by senior physicians in both training and operating environments. In recent years concern regarding poor surgical dexterity [1] and the broader use of minimally invasive interventions has promoted efforts to better characterize operator performance [2] and improve the effectiveness and efficacy of training [3]. Analytical approaches, such as task partitioning [4], kinematics analysis [5], off-line "data mining" to train hidden Markov models [6] or support vector machines [7], and many others have been developed. These techniques may now be used to measure the potential value of new interventional systems, for example determining the value of augmented reality displays to assist an endoscopist or surgeon in performing procedures [ref withheld] and to elucidate features which are most "helpful" to guide further development.

⋆ This work has been funded by NIH/NCI under award 2 R42 CA115112-02A2 and the Center for Integration of Medicine and Innovative Technology (CIMIT), Boston, MA.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 363–372, 2010.

Here, we consider colonoscopy, wherein a highly flexible endoscope (colonoscope) is inserted into the colon. The colon is elastic and deforms under the force applied by the colonoscope. Our initial study suggests that using just the location and kinematics of the distal dip of the colonoscope is not sufficient to classify the operator's performance. In addition, it was also observed that users had significantly different performance in each section of the colon. Typically, it is most difficult to maneuver the scope in the transverse and ascending colon region since the length of the colonoscope inserted is large, resulting in greater flexing of the scope. Kinematics based metrics do not identify these differences or highlight the gestures required to manipulate the scope within the colon. Here we develop and evaluate a probabilistic approach based on the Hidden Markov Model (HMM) to classify the operator performance. Also, we establish a model of expert performance to analyze and classify the ability of colonoscopy trainees. Other investigators, [8], [6], [9], [10], have used HMM techniques to analyze operator performance. We build on these studies to characterize colonoscopy, including the use of flexible instruments, identifying the operator's performance in each segment of the colon, and specifying the gestures for performing colonoscopy at the expert level. Characterizing the expertise of a user would be useful in developing curricula and simulators to train operators to smoothly guide the colonoscope with minimal discomfort to the patient.

2 Experimental Setup

The experimental setup (Figure 1) consists of a colon model (CM-1, Olympus, Tokyo, Japan), which closely mimics the human colon and includes the ascending, descending and transverse colon. The model is loosely tethered to the back support, allowing it to flex and stretch, as observed in an actual procedure. The model is draped with a cloth to prevent the user from observing the location of the scope inside the model. A pediatric colonoscope (PCF-Q180AL, Olympus, Tokyo, Japan) is equipped with four electromagnetic 6-DOF position sensors ("Microbird" sensors from Ascension Technologies Corp. (ATC), Burlington, VT). The sensors are placed at 0cm, 10cm, 30cm and 55cm from the distal end. Sensor 1 and sensor 2 are placed to record the angulation of the distal end of the scope in 2-DOF about the y and z-axis. The position of sensors 3 and 4 arechosen such that these sensors are approximately in the recto-sigmoid junction when the distal end of the scope is in the traverse colon region, thereby permitting the detection of flexing and looping of the scope. The ATC electromagnetic system is connected to an Intel Quad Core 2GHz computer with 4GB RAM. The position readings are logged at a sampling rate of 67 Hz using MATLAB Simulink.

Four attending endoscopists who have performed more than 2000 colonoscopies ("Experts") and 9 gastroenterology fellows (3 first-year, 3 second-year, 3 third-year) who have performed less than 500 colonoscopies were selected to perform a colonoscopy. Kinematics data consisting of the position and orientation of the four sensors, and time were recorded from the instant of insertion of the scope into the anus to the instant when the terminal ileum was intubated. The

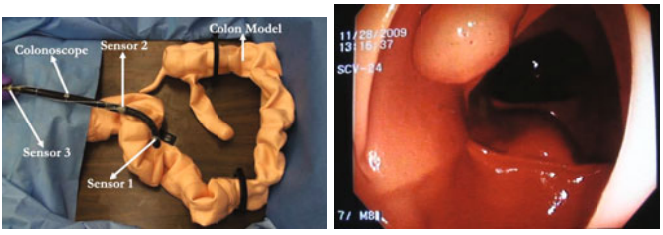


Fig. 1. (a) Colon model and colonoscope showing the position of sensor 1, 2 and 3. Sensor 4 is out of the field. (b) Inner view of the model showing a realistic modeling of the human colon.

trajectories recorded for two expert clinicians were selected to train the expert HMM model. In addition, an experienced resident with considerable training on the colon simulator was chosen to perform 5 insertions and retractions of the colonoscope into each segment of the model. During this experiment, the distal tip of the colonoscope in the colon model was tracked visually during the insertion. These trajectories were used to train three models corresponding to the motion of the scope in the ascending, descending and transverse colon. The "roll" and the "angulation" of the colonoscope in 2-DOF were computed from the measurements of the distal two sensors: these were used as features to be recognized by the HMM.

3 Hidden Markov Model

Our HMM analysis of colonoscopy is based on the approach and notation of Rabiner [11]. HMM analysis to quantify surgical expertise is suitable when the measurements obtained from the sensors on the colonoscope are statistically correlated to the measurements obtained from other subjects with a similar level of expertise. The parameters of the HMM model are defined as follows:

- The HMM is assumed to have N states. The transition probability between state i to j is given by

$$a_{ij} = P(q_{t+1} = S_j | q_t = S_i) \text{ and } A = \{a_{ij}\} \tag{1}$$

- Each state also has M possible observation symbols O_t. The probability of observing a particular symbol O_t in state j is

$$b_j = P(O_t | q_t = S_j) \text{ and } B = \{b_j\} \tag{2}$$

- Also a state prior π_i is defined, which is initial probability of beginning in S_i

In short the HMM can be represented as $\lambda = (A, B, \pi)$. Therefore, to completely define the HMM, we should define the number of states N, the observation symbols M per state and the probability measure λ. It was observed in our case, that $N = 8$ and $M = 16$ provided the best results. A larger value of N or M led

to greater complexity of the model and insufficient training while a smaller value of N and M led to an extremely simple model without capturing the variations in the surgical gestures. The model λ is trained according to the trajectories obtained from the expert clinicians.

Short-time Fourier Transform
The observation sequence is created from the trajectory of the four sensors. First, the time-domain trajectories are converted into the corresponding Fourier transforms to extract the important information from the trajectories. The Fourier transform is invariant to rotation of the trajectory, preserves the information in the signals and can be computed efficiently. However, the Fourier transforms lack the temporal localization of the frequencies. Therefore, we use the Short-Time Fourier Transform (STFT) in short time periods and obtain a feature vector corresponding to each time window [6],[12]. The STFT is computed as,

$$STFT_x^{\gamma} = \int_{\tau} [x(\tau)\gamma(\tau - t')]e^{-j2\pi\tau} d\tau \tag{3}$$

where $\gamma(\tau - t')$ is the sampling window of the trajectory. The Fourier transform in each sampling window is computed by the Fast Fourier Transform (FFT) algorithm. Information loss is minimized by overlapping the STFT windows.

The trajectories to be recognized by the HMM are the position of the four sensors in 3-DOF Cartesian space. In each sampling window, the STFT for a single DOF trajectory of a sensor consists of the magnitude of N discrete frequency contributions. Therefore, the entire feature space corresponding to the four sensors in 3-DOF is a 12 N-tuple. In addition, we have also independently trained three HMMs corresponding to the roll of the colonoscope and 2 angulations of the distal end of the colonoscope. For training each of these HMMs, the feature map is a 1 N-tuple.

Vector Quantization
Since the HMM structure considered in this paper is discrete, we convert the 12N or N tuple vector into a single discrete observation symbol using the k-means clustering algorithm [13]. Consider that there are p 12N tuples over the entire duration of the trajectory for training the expert model. The k-means algorithm partitions the p vectors into L sets so as to minimize the within-cluster sum of squares. Here L is the size of the codebook and has been chosen as 16. The discrete observation symbol is the index of the codebook vector closest to the given 12N tuple vector, i.e., the cluster in which the vector belongs.

HMM training
Having generated the discrete observation symbols corresponding to the trajectory of the expert clinicians, the observation sequence is provided to the HMM network to obtain the updated model $\hat{\lambda}$. The parameters of the models are estimated by maximizing the auxiliary function

$$Q(\lambda, \bar{\lambda}) = \sum_{Q} P(Q|O, \lambda) \log[P(O, Q|\bar{\lambda})] \tag{4}$$

This optimization problem is solved iteratively by the Baum-Welch method [11]. Seven different HMM models were trained corresponding to the Expert, Ascending Colon, Transverse Colon, Descending Colon, Roll, Angulation about y-axis and Angulation about z-axis. The Expert model was trained for the trajectories of two expert clinicians while the Ascending Colon, Transverse Colon, Descending Colon HMM models were trained on the trajectories executed by a highly experienced third year resident with several hours of practice on the colon model. The "Roll" and two "Angulation" models have been trained on the roll and angulation trajectories computed from the clinician's orientation trajectories of the distal sensors.

HMM Prediction
Once the HMMs have been trained, the next step is to measure whether the HMM classifies the expertise of the operators. That is, we evaluate the likelihood that a particular HMM describes the observation sequence. Input data includes the position trajectories of the four sensors, and the roll and 2-DOF angulation of the distal end of the scope. The probability of predicting the observation sequence given the HMM model is computed inductively using the forward-backward algorithm:

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i) \tag{5}$$

where $\alpha_T(i) = P(O_1 O_2 ... O_T, q_t = S_i|\lambda)$ is the forward variable. The reader is referred to [11] for greater details.

4 Experimental Results

Thirteen GI endoscopists performed colonoscopy in the colon model. The time taken to reach the terminal ileum from the anus ranged from 82 seconds to 1065 seconds. The position measurements from the four electromagnetic sensors were logged continuously, as shown in Figure 2. Note that the orientations of the four sensors vary considerably as the colonoscope moves through the different regions of the colon (Figure 3 (a)). A number of kinematic metrics were computed and are shown in Table 1. The position trajectories of the four sensors were provided as input to the trained Expert HMM model. In addition, the orientation of the four sensors with respect to the electromagnetic transmitter were also logged. Based on the orientation of sensor 2, the position and orientation trajectories were segmented into descending, transverse and ascending colon. These trajectories were provided as input to the trained HMM models corresponding to the Descending, Transverse and Ascending colon. The distal end of the colonoscope is capable of bending in 2-DOF about the y and z axis. From the orientation of sensor 1 and sensor 2, the angulation of the colonoscope in 2-DOF was computed and is shown in Figure 4(a). In addition, the roll of the colonoscope was calculated by measuring the roll of the local frame of the sensor at any time point with respect to the initial frame of reference, as shown in Figure 4 (b).

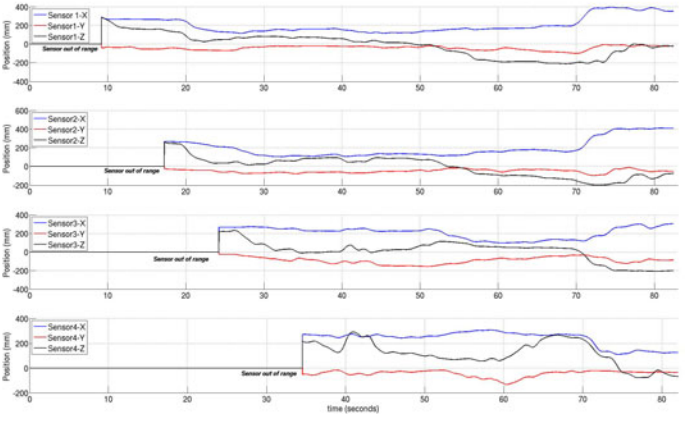


Fig. 2. Trajectory of the four electromagnetic sensors

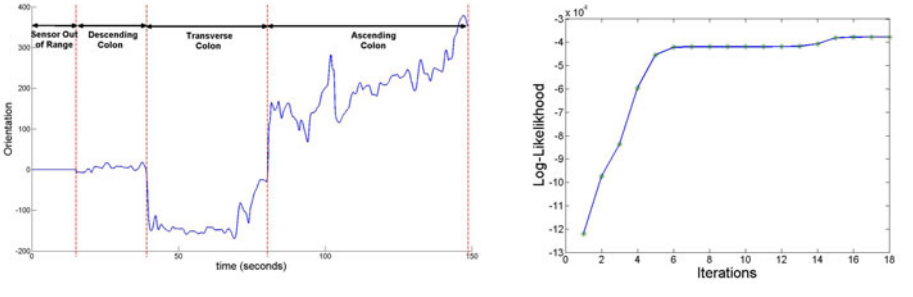


Fig. 3. (a) Variation in orientation of the sensor 2 in the descending, transverse and ascending colon. (b) Log-likehood as a function of iteration during training.

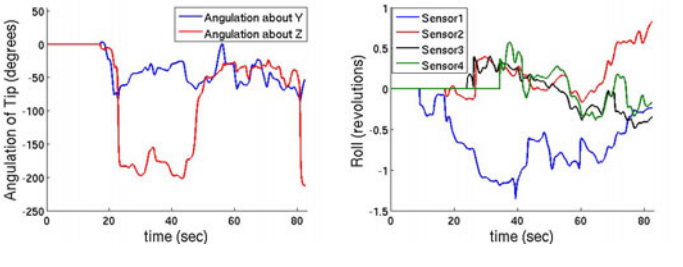


Fig. 4. (a) Graph showing the angulation of the distal end of the scope (b) Graph showing the roll of the four sensors

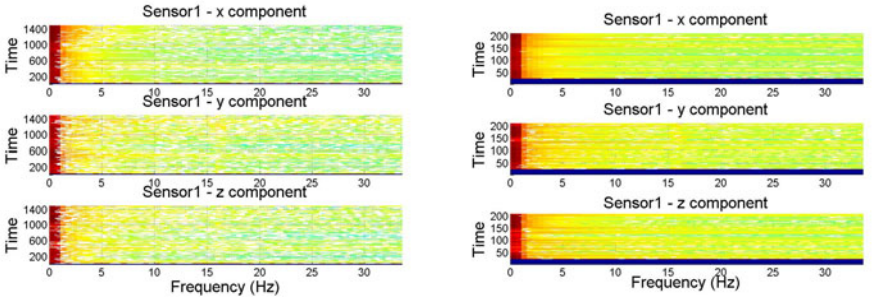


Fig. 5. (a) Spectral distribution of the motion of the colonoscope manipulated by a novice (b) Spectral distribution of the motion of the colonoscope manipulated by an expert

The spectral analysis of the position trajectory of a novice and expert are shown in Figure 5. The STFT of the position trajectory were converted to discrete observation symbols and provided to the different HMM models to first train and then predict the performance of the user. During training, only 2 expert trajectories were utilized to update the model parameters corresponding to the Expert HMM. The number of iterations required for complete training of the HMM was 18 and the log-likehood of observing the training sequence as a function of the iteration is shown in Figure 3 (b). The trained HMM classifies the performance of an operator based on the manipulation of the colonoscope by the user. The result of the prediction from the Expert model is shown in Figure 6 (a). In addition, the segmented descending, ascending and transverse colon position trajectories were provided to the corresponding trained HMM models. These models provide insight into the expertise of the user in manipulating the colonoscope in the corresponding regions of the colon. The result of the prediction of the user's performance in the three segments of the colon is shown in Figure 6 (a). In addition, the roll and angulation trajectories were provided as

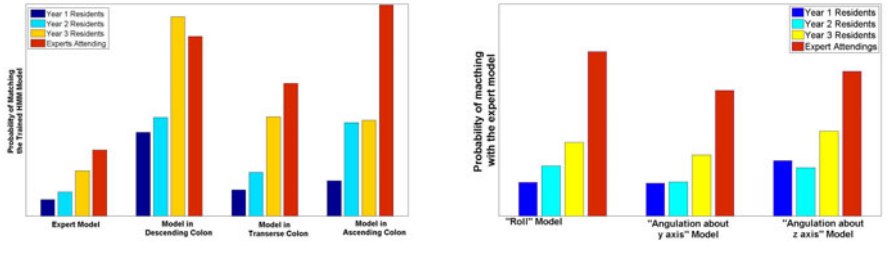


Fig. 6. (a) Evaluation of surgical performance of the users in manipulating the colonoscope compared to the expert model. The bar graph also show the performance of the users in different region of the colon. (b) Evaluation of the roll and angulation gestures compared to the expert model.

input to the HMM trained for identifying the gestures corresponding to the rotation and bending of the distal end of the colonoscope. The results corresponding to the HMM prediction of these gestures are shown in Figure 6(b).

5 Discussion

We conclude from Table 1 that the time taken for completion of the procedure is far less for an expert clinician than for the fellows. In addition, the average path length of the four sensors for the expert group is less compared with that of the first, second and third year fellows. However, none of the simple kinematics parameters shows a significant difference among the four groups of users. Further, path length and time are not ideal kinematics parameters since in an actual colonoscopy procedure, the operator could take time or move the scope locally to study a particular feature or lesion. That is, path length and time are not only a function of the expertise of the clinician but also of the complexity of the procedure. In addition, the kinematics metrics provide a single global metric to quantify the performance of the clinician, which is not sufficient to analyze the trajectories in detail. For example, it would be useful to analyze the performance of the user in various regions. Typically, in conventional surgical training, the expert surgeon is considered as the gold standard of performance and the novices are trained to follow the expert's movements. This method of training may be captured by the HMM by comparing the novice's performance to the expert.

It is observed in Figure 5 that the experts have a larger frequency component compared to novices during the entire duration of the procedure. This is contradictory to the findings in [6] wherein it is observed that the novices have a larger frequency component in manipulating a laparoscope (rigid surgical tool) compared to the experts. It is our hypothesis that due to the flexibility of the colonoscope and the elastic nature of the colon, the high frequency component of the motion of the colonoscope is dampened. In addition, the experts are observed to manipulate the distal end of the colonoscope with higher velocities (as suggested by Table 1), resulting in higher frequency components in the STFT. This may indicate that the approach adopted for training clinicians to perform laparoscopic surgery cannot be applied for endoscopy-based procedures.

From Figure 6 (a), it can be seen that the performance of the 13 subjects can be clearly classified by the Expert trained HMM model based on their known expertise. The figure shows that the first year novices are less likely to achieve

Table 1. Metrics for evaluating clinician's performance

	Time sec	Pathlength m	Flexing m	Av.Vel. mm/sec	Av.Accel. mm/sec^2	Angulation Y degrees	Angulation Z degrees	Roll rev.
1^{st} Yr	715.4	10.92	5.33	0.83	0.65	39.7	95.7	0.21
2^{nd} Yr	288.2	7.56	3.17	0.98	0.91	44.4	105.8	0.26
3^{rd} Yr	274.9	5.16	1.53	1.15	0.63	40.5	96.8	0.16
Expert	150.1	3.31	1.75	1.29	0.99	42.0	95.3	0.22

the "Expert" performance compared to other groups of users. It can also be seen that the performance of the first and second year residents is significantly less probable to match the expert performance in all three regions of the colon. However, the third year residents show comparable performance to the experts in the descending colon region. A likely explanation is that the descending colon is closest to the anus and therefore, the length of the colonoscope inserted into the model is small, resulting in less flexing in the colonoscope and easier manipulation. However, once the colonoscope enters the ascending and descending regions of the colon, the insertion becomes more difficult due to flexing and excessive curvature in the scope. This can be noticed in the performance of the residents compared to that of the experts (Figure 6 (a)). Figure 6 (b) shows the comparison of the gestures (roll and angulation of the distal end of the colonoscope) among the four groups of users. It is observed that the gestures performed by the residents are less likely to match "Expert" performance.

6 Conclusion

We have developed a Hidden Markov Model (HMM) to quantify the performance of a clinician performing colonoscopy using a realistic physical colon model. In addition, we have also analyzed the motion of the scope in each segment of the colon to identify the degree of expertise of manipulating the scope in the ascending, descending and transverse colon. We have shown that the HMM approach robustly classifies the expertise of the users based on their experience. In addition, the roll and angulation gestures are significantly different for the four groups of users and are clustered based on the expertise of the clinicians. This may provide a useful training tool to characterize the expertise of a physician in training. Further work is underway in validating the results of this work in human subjects.

References

1. Darzi, A., Smith, S., Taffinder, N.: Assessing operative skill: Need to become more objective. British Medical Journal 318, 887–888 (1999)
2. Satava, R., Cuschieri, A., Hamdorf, J.: Metrics for objective assessment: Preliminary results of the surgical skills workshop. Surgical Endoscopy 17, 220–226 (2003)
3. Peters, J., Fried, G., Swanstrom, L., Soper, N., Silin, L., Schirmer, B., Hoffman, K., et al.: Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. Surgery 135, 21–27 (2004)
4. Heinrichs, W., Srivastava, S., Montgomery, K., Dev, P.: The fundamental manipulations of surgery: A structured vocabulary for designing surgical curricula and simulators. J. Amer. Assoc. of Gynecologic Laparoscopists 11, 450–456 (2004)
5. Dosis, D., Aggarwal, R., Bello, F., Moorthy, K., Munz, Y., Gillies, D., Darzi, A.: Synchronized video and motion analysis of the assessment of procedures in the oeprating theater. Arch. Surg. 140, 293–299 (2005)

6. Megali, G., Sinigaglia, S., Tonet, O., Dario, P.: Modelling and evaluation of surgical performance using hidden markov models. IEEE Transactions on BioMedical Engineering 53, 1911–1919 (2006)
7. Allen, B., Nistor, V., Dutson, E., Carman, G., Lewis, C., Faloutsos, P.: Support vector machines improve the accuracy of performance evaluation of laparoscopic training tasks. Surgical Endoscopy, 1–14 (2009)
8. Blum, T., Padoy, N., Feubner, H., Navab, N.: Modeling and online recognition of surgical phases using hidden markov models. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 627–635. Springer, Heidelberg (2008)
9. Rosen, J., Brown, J., Chang, L., Sinanan, M., Hannaford, B.: Generalized approach for modeling minimally invasive surgery as a stochastic process using a discrete markov model. IEEE Trans. on Biomed. Eng. 53, 399–413 (2006)
10. Leong, J.J.H., Nicolaou, M., Atallah, L., Mylonas, G.P., Darzi, A.W., Yang, G.Z.: HMM assessment of quality of movement trajectory in laparoscopic surgery. Computer Aided Surgery 12, 335–346 (2007)
11. Rabiner, L.: A tutorial on hidden markov models and selected applications in speech recognition. Proceeding of the IEEE 77, 257–286 (1989)
12. Hannaford, B., Lee, P.: Hidden Markov Model analysis of force/torque information in telemanipulation. The International Journal of Robotics Research 10, 528–539 (1991)
13. Likas, A., Vlassisb, N., Verbeekb, J.J.: The global k-means clustering algorithm. Pattern Recognition 36, 451–461 (2003)

A Robust Mosaicing Method with Super-Resolution for Optical Medical Images

Mingxing Hu[1], Graeme Penney[2], Daniel Rueckert[3], Philip Edwards[4], Fernando Bello[4], Michael Figl[3], Roberto Casula[5], Yigang Cen[6], Jie Liu[7], Zhenjiang Miao[6], and David Hawkes[1]

[1] Centre for Medical Image Computing, University College London,
[2] Department of Imaging Sciences, King's College London
[3] Department of Computing, Imperial College
[4] Department of Surgical Oncology and Technology, Imperial College,
[5] Cardiothoracic Surgery, St. Mary's Hospital, London, UK
[6] Institute of Information Science, Beijing Jiaotong University, Beijing, China
[7] Department of Biomedical Engineering, Beijing Jiaotong University, Beijing, China
{mingxing.hu,d.hawkes}@ucl.ac.uk, graeme.penney@kcl.ac.uk,
{d.rueckert,eddie.edwards,f.bello,m.figl}@imperial.ac.uk,
roberto.casula@st-mary.nhs.uk, {ygcen,jieliu,zjmiao}@bjtu.edu.cn

Abstract. Constructing a mosaicing image with a broader field-of-view has become an important topic in image guided diagnosis and treatment. In this paper, we present a robust feature-based method for video mosaicing with super-resolution for optical medical images. Firstly, outliers involved in the feature dataset are removed using trilinear constraints and iterative bundle adjustment, then a minimal cost graph path is built for mosaicing using topology inference. Finally, a mosaicing image with super-resolution is created by way of maximum a posterior (MAP) estimation and selective initialization. The proposed method has been tested with both endoscopic images from totally endoscopic coronary artery bypass surgery and fibered confocal microscopy images. The results showed our method performs better than previously reported methods in terms of accuracy and robustness to deformation and artefacts.

1 Introduction

Today optical medical images from endoscopy and microscopy have been widely used in diagnosis, screening and treatment in a large variety of medical applications. For example, video endoscopic system is used to detect, localize or target the biopsy sites or some visible lesions of the interior surface of an organ. Fibered confocal microscopy (FCM) system on the other hand can provide cellular structure information of the observed tissue, which make the *in-vivo* and *in-situ* exploration of the living organs possible. By combining with other preoperative medical information, e.g., MRI or CT scan data, the combination of endoscopy with microscopy will become a very powerful tool for image guided intervention for diagnosis and surgery. However, both images acquired from endoscopes and microscopes suffer from a fundamental problem – a

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 373–382, 2010.

narrow field-of-view. As a result, the limited vision causes great difficulty for the clinician to collect the visual information and be aware of the peripheral sites.

A common solution to this problem is to build a 2D mosaicing image and a lot of research work has been done in the medical imaging community [1-8]. These methods can be roughly divided into two categories: (1) Similarity-based methods. These compute the transformation and register the images together by optimizing the similarity measure. For example, Vercauteren *et al.* proposed a mosaicing method for confocal microscopic images [1]. Pairwise registration between images was carried out using similarity-based registration and then a globally consistent result was achieved by applying statistics for Riemannian manifolds. Miranda-Luna *et al.* also used the mutual information-based similarity measure in the registration process but a stochastic gradient optimization was implemented and the correction of the endoscopic distortion was also considered in the whole framework [2]. However, similarity-based methods usually require large overlaps to deliver a reliable result. (2) Feature-based methods. These use the geometric constraints to align the images using a sparse set of image correspondences. For example, in [3], the branching and crossover points of the blood vessel were used as the features in retinal images, and a 12-parameter, quadratic transformation model was created, followed by a hierarchical estimation to register pairs of images. Cattin *et al.* also proposed a retina mosaicing method but used SURF descriptors [4] to obtain good image correspondences and presented a scheme to determine the blending masks of arbitrarily overlapping images for multi-band blending [5]. More recently, Atasoy *et al.* proposed an automatic fibroscopic video mosaicing method [6], which is similar to the algorithm developed by Brown and Lowe [9]. They considered the characteristic of fibroscope such as lens distortion and high-frequency fibered optic facet pattern. However, a common problem for feature-based methods is the outlier, which is caused by bad localization or false matching. Outliers can distort the estimation to such a great extent that the fitted parameters may become essentially random.

In this paper, we propose a robust feature-based mosaicing method with super-resolution for optical medical images. Our contributions are as follows: (1) The proposed method is robust to outlier and even to artefact at a certain level. Trilinear constraints among three images are used to remove the potential outliers from the image feature dataset. Then iterative bundle adjustment is applied to detect the remaining outliers involved in the mosaicing refinement based on a Maximum Likelihood Estimate (MLE). (2) The proposed method can deliver a final mosaicing image with super-resolution. Maximum a posterior (MAP) estimation was applied to generate the super-resolution image with a good initialization from a selective averaging result. (3) The proposed method can be applied to build mosaicing image not only for endoscopic video but also for microscopic (e.g., fibered confocal microscopy) image sequence. The mosaicing image with wider field-of-view will provide more useful 2D information for image guided diagnosis and surgery.

2 Method

2.1 Outlier Removal with Trilinear Constraints

In this paper, a Lucas-Kanade (LK) tracker based on optical flow was applied to detect and track some features from the image sequence [10]. Readers can refer to [11] for

more details. However, outliers are likely o be involved in the feature dataset and they can be in gross disagreement with a specific postulated model. So we use trilinear constraints to detect and remove the outliers.

The trilinear constraints across the three views, with correspondences $\mathbf{x}=(x, y, 1)^T$, $\mathbf{x}'=(x', y', 1)^T$, $\mathbf{x}''=(x'', y'', 1)^T$, can be compactly expressed in terms of the trifocal tensor, $\mathbf{T}_i^{jk}$, which is a $3\times3\times3$ matrix [12]

$$\mathbf{T}_i^{jk} = \mathbf{v}'^j\mathbf{b}_i^k - \mathbf{v}''^k\mathbf{a}_i^j, \quad i,\ j,\ k=1,\ 2,\ 3 \tag{1}$$

where $\mathbf{a}$, $\mathbf{b}$, $\mathbf{v}'$ and $\mathbf{v}''$ are from projective matrices of the three views, $\mathbf{P}=[\mathbf{I}|\mathbf{0}]$, $\mathbf{P}'=[\mathbf{A}|\mathbf{v}']$ and $\mathbf{P}''=[\mathbf{B}|\mathbf{v}'']$. It has been shown that the minimal number of independent equations for trilinear constraints is four. Thus Eq. (1) can be rewritten as

$$\begin{cases} x''\mathbf{T}_i^{13}\mathbf{x}^i - x''x'\mathbf{T}_i^{33}\mathbf{x}^i + x'\mathbf{T}_i^{31}\mathbf{x}^i - \mathbf{T}_i^{11}\mathbf{x}^i = 0 \\ y''\mathbf{T}_i^{13}\mathbf{x}^i - y''x'\mathbf{T}_i^{33}\mathbf{x}^i + x'\mathbf{T}_i^{32}\mathbf{x}^i - \mathbf{T}_i^{12}\mathbf{x}^i = 0 \\ x''\mathbf{T}_i^{23}\mathbf{x}^i - x''y'\mathbf{T}_i^{33}\mathbf{x}^i + y'\mathbf{T}_i^{31}\mathbf{x}^i - \mathbf{T}_i^{21}\mathbf{x}^i = 0 \\ y''\mathbf{T}_i^{23}\mathbf{x}^i - y''y'\mathbf{T}_i^{33}\mathbf{x}^i + y'\mathbf{T}_i^{32}\mathbf{x}^i - \mathbf{T}_i^{22}\mathbf{x}^i = 0 \end{cases} \tag{2}$$

Since seven point correspondences uniquely determine (up to scale) the tensor $\mathbf{T}_i^{jk}$ [12], we use the seven-point method to compute a possible solution and employ the RANSAC strategy [13] to detect the outliers based on the geometric error.

$$E = \sum_{l=1}^{n} E_l = \sum_{l=1}^{n} d\left(\mathbf{x}_l^i, \hat{\mathbf{x}}_l^i\right)^2 + d\left(\mathbf{x}_l^j, \hat{\mathbf{x}}_l^j\right)^2 + d\left(\mathbf{x}_l^k, \hat{\mathbf{x}}_l^k\right)^2 \tag{3}$$

This error measures the sum-of-squares of the geometric distances between the correspondences $\mathbf{x}_l' \leftrightarrow \mathbf{x}_l'' \leftrightarrow \mathbf{x}_l'''$ and the corrected points $\hat{\mathbf{x}}_l' \leftrightarrow \hat{\mathbf{x}}_l'' \leftrightarrow \hat{\mathbf{x}}_l'''$, with the latter obeying the trilinear constraint for the estimated tensor $\mathbf{T}_i^{jk}$. Thus, given three images with overlap, we can estimate the trifocal tensor and use the above error measure to detect outliers accordingly. In order to process an entire image sequence, we compute the tensor among nearby images, $(i, i+1, i+2)$ and $(i, i+1, i+3)$. A feature point is removed only if it is determined as an outlier more than 50% of the times of some independent trilinear constraint testing.

2.2 Graph-Based Registration

After the outlier removal, the homography between consecutive images can be computed to align the two images. Ideally, after the alignment of all consecutive images, we can chain all the images together and warp them onto a reference plane. However, the misalignment error usually accumulates by concatenating homographies. This is especially evident when the camera goes back to the scene previously seen in a long image sequence.

So in our method, we build a graph path to align all the images based on the topology inference [14, 15]. The graph is composed of vertices (the image frames) and arcs (the mapping between two overlapping images). Firstly the arcs connecting the consecutive images are added to the graph. Then more arcs will be added incrementally, if they satisfy the following constraints:

(1) Having a sufficient overlap $\alpha^{i,j}$, the normalized distance between centroids

$$\alpha^{i,j} = \max\left(0, \left|\mathbf{x}_c^i - \mathbf{x}_c^j\right| - \left|d^i - d^j\right|/2\right) / \min\left(d^i, d^j\right)$$

where $\mathbf{x}_c^i$, $\mathbf{x}_c^j$, d^i, d^j are the centroids and diameters of the projection onto the mosaicing image of frame I^i and I^j, respectively.

(2) Having sufficient image correspondences, $n_{inlier}^{i,j} \geq \delta_{\min\_matching}$

In contrast to [14, 15], we also make sure there are enough inliers between images to deliver a reliable homography since the features usually do not spread evenly around the images and some of them have been removed as outliers using the strict trilinear constraints.

(3) Creating a material shortest path between two vertices: $\beta^{i,j} = \alpha^{i,j} / \Delta^{i,j}$.

$\Delta^{i,j}$ is the cost of the shortest path between image i and j [14], in the graph with weights $\alpha^{i,j}$ on the edges. The higher $\beta^{i,j}$ is, the less influence it is likely to have a pair (i, j) on the alignment.

Here a simple greedy algorithm is used to select good arcs that need to be added to the graph. The complete algorithm is shown in Fig. 2. Finally the frame-reference transformation $\mathbf{H}^{i,r}$ (r is the reference frame), for each frame i, is computed by using the graph to chain homographies along the minimal cost path.

2.3 Iterative Refinement

The bundle adjustment (BA) we used is different from the ones addressed in [11, 16], in which focal length and motion parameters were also estimated in the minimization. In this paper, BA was performed to find the best homography set $\{\mathbf{H}^{i,r}\}$, $i = 1, \cdots, m$, that minimize the misalignment error.

$$\min_{\{\mathbf{H}\}} \sum_{(i,j)} \sum_{k=1}^{m} \left\| \mathbf{x}_k^i - \left(\mathbf{H}^{r,i}\right)^{-1} \mathbf{H}^{r,j} \mathbf{x}_k^j \right\| \tag{4}$$

In each iteration, when BA refinement is finished, we compute the average projection error of each feature

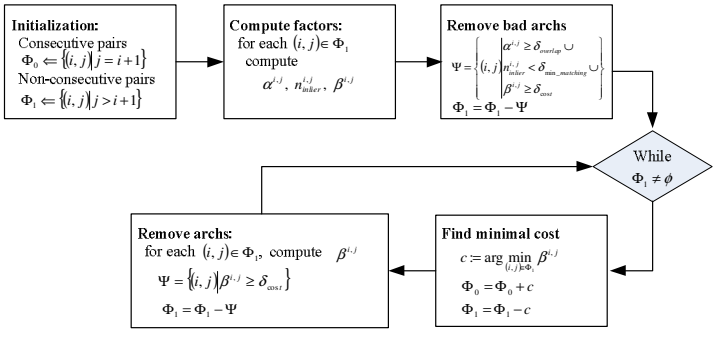


Fig. 1. The complete algorithm for graph-based registration for image mosaicing

$$D_k = \frac{1}{s}\sum_{(i,j)} \left\| \mathbf{x}_k^i - \left(\mathbf{H}^{r,i}\right)^{-1} \mathbf{H}^{r,j} \mathbf{x}_k^j \right\| \tag{5}$$

where s is the number of image pair (i, j) in which feature $\mathbf{x}_k$ is used to calculate the homography. The potential outlier is detected if the reprojection error is larger than the threshold $\gamma = 1.96\sigma$. The standard deviation σ can be found as a maximum likelihood estimate using the median $\sigma = 1.4828\left(1+5/(n-8)\right)\sqrt{med_i |D_i|}$. The iterative BA will continue until no outlier can be detected from the feature dataset.

2.4 Super-Resolution Mosaicing

In order to generate a mosaicing image with super-resolution, information from images of different viewpoints is combined together based on Maximum a Posterior (MAP) technique [17]. The geometric mapping between the original image and super-resolution image can be expressed as $\hat{\mathbf{I}}^i = \mathbf{G}^i \hat{\mathbf{S}}$. $\hat{\mathbf{I}}^i$ denotes the predicted intensities of the original image $\mathbf{I}^i$ and $\hat{\mathbf{S}}$ denotes the intensities of the super-resolution image. The geometric matrix $\mathbf{G}^i$ is computed from homographies and interpolation via area-sampling [17]. If we stack all these images into one matrix, we have $\hat{\mathbf{I}} = \mathbf{G}\hat{\mathbf{S}}$. $\hat{\mathbf{I}}$ and $\mathbf{G}$ are the stacks of predicted original image $\mathbf{I}^i$ and geometric matrices $\mathbf{G}^i$, respectively. Usually this problem can be solved by maximum likelihood estimation to minimize the residual between the original image $\mathbf{I}$ and the predicted image $\hat{\mathbf{I}}$

$$R = \left\| \mathbf{G}\hat{\mathbf{S}} - \mathbf{I} \right\|^2$$

However, this solution is very sensitive to image noise and misregistration caused by the homography estimation. So MAP estimation can be used to improve the robustness by bringing in a regularizing term to penalty high gradients in the estimated super-resolution image. So the objective aims to find a value of $\mathbf{S}$ which maximizes $\Pr(\mathbf{S}|\mathbf{I}) = \Pr(\mathbf{I}|\mathbf{S})\Pr(\mathbf{S})$ and the cost function turns into

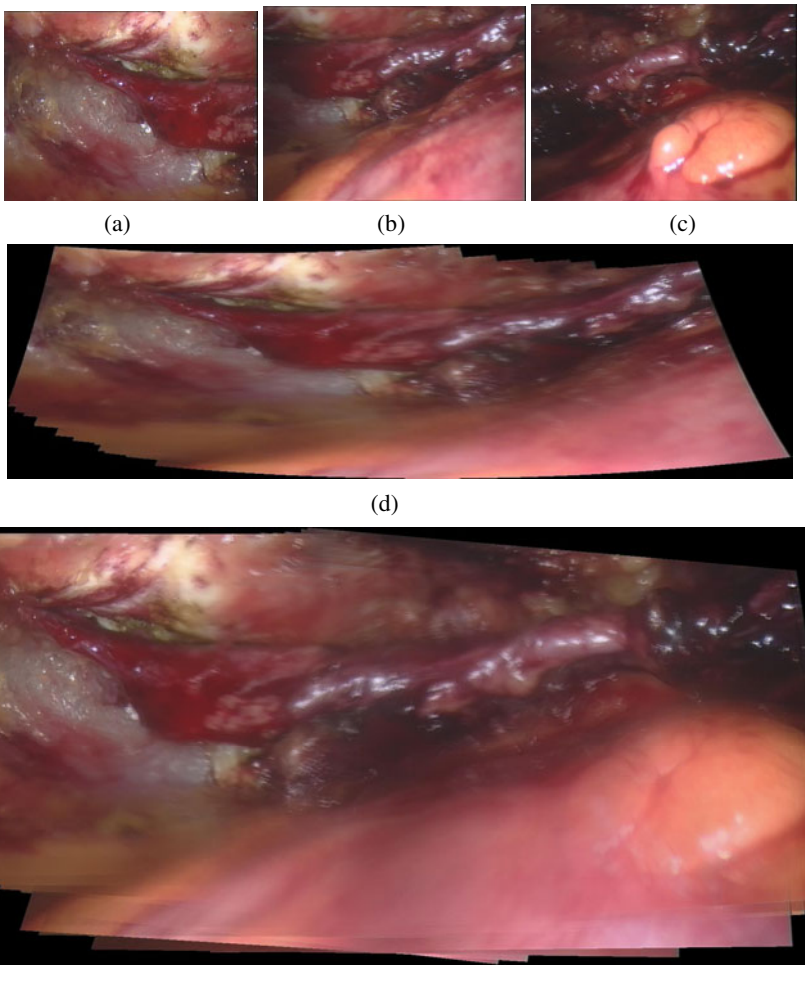
$$R = \left\| \mathbf{G}\hat{\mathbf{S}} - \mathbf{I} \right\|^2 + \lambda \left\| \Lambda\left(\hat{\mathbf{S}} - \overline{\mathbf{S}}\right) \right\|^2$$

where $\overline{\mathbf{S}}$ is the initial estimate of the super-resolution image $\mathbf{S}$ and $\Lambda = diag\left(\left|\nabla \overline{\mathbf{S}}\right|\right)^{-1}$, the reciprocal of the image gradient. Scale λ usually ranges from 0.1 to 1.0 and in our experiments it is set to 0.5. Since a good initial estimate $\overline{\mathbf{S}}$ is very important to the success of MAP estimation, the original images are first projected to the reference image plane via the minimal cost graph path and then the selective averaging process is used to combine the information from different images together. That is, for each pixel, intensities around the pixel (3×3 or 5×5) are also taken into account to remove outlier-pixels from some image with very different similarities. This can help us deal with the deformation caused by the internal organ or soft tissue.

3 Experimental Results

3.1 Experiment with Endoscopic Images from TECAB Surgery

The *da Vinci*™ robotic surgical system (Intuitive Surgical, Inc., Sunnyvale, CA, USA) was used to obtain images of the heart surface. The endoscopic images were digitized at 25 frames per second (fps) using a frame grabber (LFG4 PCI64, Active Silicon, Uxbridge, U.K.). 150 images were captured from the left camera but we used only 30 frames (every 5 frames from the sequence) for the mosaicing. Our aim is to create a



(a) (b) (c)

(d)

(e)

Fig. 2. The experimental results using the endoscopic images from TECAB surgery. (a), (b) and (c) show the first, middle and last images of the sequence, respectively. (d) and (e) display the mosaicing results of Brown's and the proposed methods, respectively.

mosaicing image which includes the majority of the structure of the targeted vessel. The main challenge is the large complicated non-rigid motion introduced by the beating heart surface, which is shown in the right bottom of Fig. 2 (b) and (c).

Fig. 2 (e) displays the mosaicing result of the proposed method. It can be seen that the whole vessel structure has been built correctly. From Table 1, we can also notice that a number of outliers was detected and removed by way of trilinear constraints (52 points), and iterative refinement (9 points), and the reprojection error was reduced from 2.36 pixels after consecutive registration to 2.12 pixels after graph-based registration, further to 1.65 pixels after iterative refinement. So the surgeon can realize the environment outside the current scene when he views a part of the vessel. More importantly, the mosaicing image can provide more information to help him to link the endoscopic video with the preoperative data from CT/MRI scan. Brown's method [11] was also tested using this image sequence and the mosaicing result was displayed in Fig. 2 (d). Note that only part of the whole vessel was constructed and the images affected seriously by the beating heart surface could not be used for Brown's method. The most probable reason for this is that the SIFT descriptor could not find enough reliable features from the images with severe deformation.

3.2 Experiment with FCM Images

A FCM sequence was acquired from *Cellvizo* System (Mauna Kea Technologies, Paris, France) when the probe moved along a sponge phantom. There were 100 frames in the sequence and we used only half of them (every other frame) for the mosaicing. The main challenges of this sequence lie in: (1) The deformation of the micro-structure in the image, as shown in Fig. 3 (b), (c) and (d) within the red circle. (2) The artefact in the image, as shown in Fig. 3 (b), (c) and (d) within the blue circle. This is caused by a very small piece of tissue attached to the probe lens, which is common in FCM biopsy.

Fig. 3 (e), (f) and (g) show the experimental results of Brown's [9], Vercauteren's [1], and the proposed method, respectively. Table 1 also displayed the number of features and reprojection error after different stages for the FCM sequence. It can be seen that our method performs best in the testing. Brown's method was misled by the SIFT features detected on the artefact, which led to the failure of whole process. Vercauteren's method yields a mosaicing result but it contained several small mosaicing pieces. Each of them covered a small part of the scene and Fig. 3 (f) showed one of them. The possible reason is that the fast probe motion and large deformation caused insufficient reliable overlap for this method to process the similarity-based registration. Our method aligned all the images correctly and covered most part of the biopsy scene, as displayed in Fig. 3 (g). The mosaicing image will help the clinician to interpret the acquired data and make quantitative and statistical analysis possible by providing a broader field-of-view.

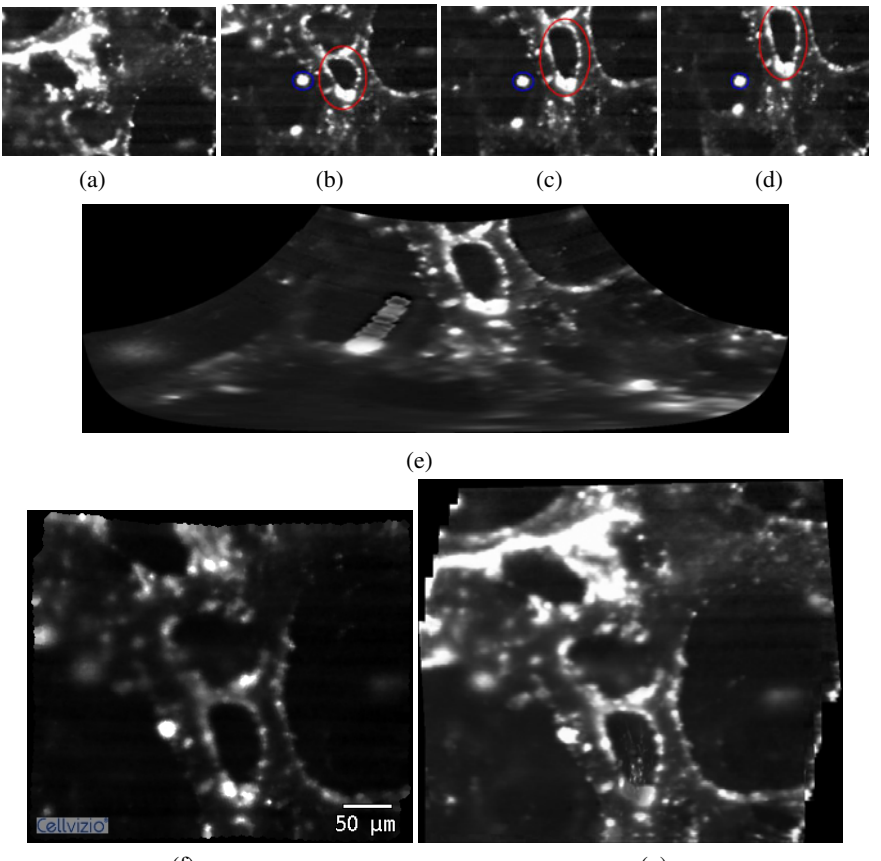


Fig. 3. The experimental results using images from fibered confocal microscopy. (a), (b), (c) and (d) show the some images from the sequence. Deformation is marked with red circle and artefact is marked with blue circle. (e), (f) and (g) display the mosaicing results of Brown's Vercauteren's and the proposed methods, respectively.

Table 1. Number of features and reprojection error after different stages of image mosaicing

	Number of features			Average reprojection error		
	Original	Removed in trilinear constraint	Removed in iterative refinement	Consecutive registration	Graph-based registration	Iterative refinement
TECAB	329	52	9	2.36	2.12	1.65
FCM	433	87	3	0.756	0.739	0.673

4 Discussion and Conclusions

In this paper, we proposed a robust video mosaicing method for optical medical images. The mosaicing image displays a much wider field-of-view of the scene and helps the clinicians realize the surrounding environment outside the current view. Experiments with TECAB endoscopic images and FCM images show that the proposed method performs better than other typical methods. It is robust to deformation in organs and soft tissues and can even deal with artefacts resulting from extraneous structure in the image.

Effort in the near future will focus on improvement of robustness to deformation and artefacts. Our long term goal is to automatically construct mosaicing image of the surgical scene, reconstruct the internal organ surfaces and register these with the pre-operative data to provide more information for image guided diagnosis and treatment.

References

1. Vercauteren, T., Perchant, A., Malandain, G., Pennec, X., Ayache, N.: Robust mosaicing with correction of motion distortions and tissue deformation for in vivo fibered microscopy. Medical Image Analysis 10(5), 673–692 (2006)
2. Miranda-Luna, R., Daul, C., Blondel, W.C.P.M., Hernandez-Mier, Y., Wolf, D., Guillemin, F.: Mosaicing of bladder endoscopic image sequences: distortion calibration and registration algorithm. IEEE Trans. on Biomed. Eng. 55, 541–553 (2008)
3. Can, A., Stewart, C.V., Roysam, B., Tanenbaum, H.L.: A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. IEEE Trans. PAMI 24, 347–364 (2002)
4. Bay, H., Tuytelaars, T., Gool, L.V.: SURF: Speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
5. Cattin, P.C., Bay, H., Van Gool, L., Szekely, G.: Retina Mosaicing Using Local Features. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 185–192. Springer, Heidelberg (2006)
6. Atasoy, S., Noonan, D., Benhimane, S., Navab, N., Yang, G.-Z.: A global approach for automatic fibroscopic video mosaicing in minimally invasive diagnosis. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 850–857. Springer, Heidelberg (2008)
7. Lerotic, M., Chung, A.J., Clark, J., Valibeik, S., Yang, G.Z.: Dynamic view expansion for enhanced navigation in Natural Orifice Transluminal Endoscopic Surgery. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 467–475. Springer, Heidelberg (2008)
8. Mountney, P., Yang, G.Z.: Dynamic view expansion for minimally invasive surgery using simultaneous localization and mapping. Proc. IEEE Eng. Med. Biol. Soc., 1184–1187 (2009)
9. Brown, M., Lowe, D.G.: Automatic panoramic image stitching using invariant features. International Journal of Computer Vision 74, 59–73 (2007)
10. Lucas, B., Kanade, T.: An Iterative Image Registration Technique with an Application to Stereo Vision. In: Proc. IJCAI, pp. 674–679 (1981)

11. Baker, S., Matthews, I.: Lucas-Kanade 20 Years On: A Unifying Framework. International Journal of Computer Vision 56(3), 221–255 (2004)
12. Shashua, A.: Algebraic functions for recognition. IEEE Trans. PAMI 17(8), 779–789 (1995)
13. Fischler, M.A., Bolles, R.C.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Comm. of the ACM 24, 381–395 (1981)
14. Sawhney, H.S., Hsu, S., Kumar, R.: Robust video mosaicing through topology inference and local to global alignment. In: Burkhardt, H., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1407, pp. 103–119. Springer, Heidelberg (1998)
15. Marzotto, R., Fusiello, A., Murino, V.: High resolution video mosaicing with global alignment. In: Proc. CVPR, pp. 692–698 (2004)
16. McLauchlan, P.F., Jaenicke, A.: Image mosaicing using sequential bundle adjustment. Image and Vision Computing 20, 751–759 (2002)
17. Capel, D.P.: Image Mosaicing and Super-Resolution, PhD thesis, Dept. of Eng. Science, Univ. of Oxford (2001)

Spectral Aggregation Based on Iterative Graph Cut for Sonographic Breast Image Segmentation

Chi-Hsuan Tsou[1], Jie-Zhi Cheng[2], Jyh-Horng Chen[1], and Chung-Ming Chen[2]

[1] Graduate Institute of Biomedical Electronic and Bioinformatics,
[2] Institute of Biomedical Engineering,
National Taiwan University, Taipei, Taiwan
d97945008@ntu.edu.tw

Abstract. In this paper, an image segmentation framework is proposed by unifying the techniques of spectral clustering and graph-cutting to address the difficult problem of breast lesion demarcation in sonography. In order to alleviate the effect of speckle noise and posterior acoustic shadows, the ROI of a sonogram is mapped to a specific eigen-space as an eigenmap by a constrained spectral clustering scheme. The eigen-mapping is boosted with the incorporation of partial grouping setting and then provide a useful preliminary aggregation based on intensity affinity. Following that, an iterative graph cut framework is carried out to identify the object of interest in the projected eigenmap. The proposed segmentation algorithm is evaluated with four sets of manual delineations on 110 breast ultrasound images. The experiment results corroborates that the boundaries derived by the proposed algorithm are comparable to manual delineations and hence can potentially provide reliable morphological information of a breast lesion.

Keywords: Breast ultrasound images, Lesion segmentation, Spectral clustering, Graph cut, Gaussian Mixture Models.

1 Introduction

Breast cancer has been reported as the second leading cause of cancer deaths for females in the world. To increase the survival rate, early detection deems to be very crucial in the war of fighting with breast cancer. In recent years, sonographic screening has been commonly adopted for the purpose of early detection and differential diagnosis of breast cancer, especially in many Asian countries. However, the interpretation of sonography is relatively difficult, and hence the identification of a breast lesion in sonography heavily depends on physician's experience. To improve this drawback, computer-aided diagnosis (CADx) schemes have recently been widely and intensively investigated for boosting the identification performance of breast lesion in sonography. In general, CADx systems involve the process of retrieving morphological and textural features from a suspected lesion for differential diagnosis of tumor malignancy [1,2]. Particularly, morphological features are shown to hold robust differentiation power for the analysis of sonographically imaged breast lesions [1,2,3].

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 383–392, 2010.

The extraction of morphological features commonly requires an image segmentation procedure to provide quantitative boundary information of a breast lesion. In this case, an effective image segmentation algorithm deems to be a critical component to in a CADx system.

Image segmentation problem for the demarcation of breast lesions in sonography is intrinsically an arduous task. The difficulty can be summarized into two aspects. First, because ultrasound images are reconstructed from signals of acoustic reflection, sonograms may be presented with shadows and speckle noise. Hence, processing of sonograms may generally encounter tissue inhomogeneous problem [4]. The second factor that complicates the task of lesion demarcation is the irregular shapes and echogenicity of breast lesions. Since any tumor can develop into any kind of shape and degree of solidity inside, rarely effective shape and echogenicity priors can be exploited to assist the task of computerized demarcation of breast lesions in sonograms. Accordingly, the development of breast lesion demarcation algorithm may need to carefully address the two mentioned problems to achieve promising performance. In the literature, prominent image segmentation methods for ultrasound images can be categorized into watershed-based approaches [5,6], deformable models [7], level-set methods [8,9], and spectral clustering [10]. Although satisfactory results had been reported in each study, the tissue inhomogeneous problem and irregular properties of sonographically imaged breast lesions were not well addressed. For examples, deformable models and level-set methods may be easily trapped into undesired solutions due to the low image quality of sonography.

To compensate the low image quality of sonography, prior knowledge of the object of interest is recommended to be incorporated in the segmentation scheme [4,11]. The prior knowledge of the object of interest can be its echogenicity distribution, shape and so on. The prior knowledge can be easily incorporated into deformable models and level-set methods. However, in the problem of breast lesion demarcation, rarely useful priors can be explored from various kinds of breast lesions. Watershed-based approaches [5,6] had been shown to be robust to speckle noise by taking the catchment basins as the basic processing units. Nevertheless, over-segmentation problem is not guaranteed to be solved by these watershed-based approaches [5,6]. Spectral clustering techniques [10,12,13] aim to segregate the image or Region of Interest (ROI) from the projected affinity subspace, which is spanned by a specific eigenvector. The spectral clustering technique was further improved to deal with clutter surrounding environment by constrained normalized cuts framework [12,13]. The constrained normalized cuts framework can map the image/ROI into analytically meaningful eigenvector domain as shown in Fig. 2(b), 2(e), and 2(h). Due to the incorporation of partial grouping bias, the effect of speckle noise and posterior acoustic shadows can be alleviated significantly. Although several issues, i.e. the determination of the cut-off threshold to delineate foreground and definition of partial grouping bias, may need to be addressed, the projected eigenmap offers a good initial partition. Note that the eigenmap we mention here is the second largest eigenvector of the constrained grouping system. Accordingly, an iterative graph cut framework [14,15] is adopted in this paper to automatically demarcate the foreground, i.e. a breast lesion, by seeking the best partition on the eigenmap.

The proposed method is constituted of two major steps. In the first step, constrained normalized cut is carried out to map the sonogram/ROI into analytic eigenvector

domain. The eigenmap is able to manifest tissues with similar echogenicity. The key implementation issues of the constrained normalized cuts [12,13] consist in the definition of the similarity measurement among pixels and the setting of partial grouping bias. The two implementation issues will be elaborated in section 2.1. The second step is devised to segregate the eigenmap into foreground (breast lesion) and background with an iterative graph cut framework. To assist the partition process, Gaussian Mixture Models (GMMs) are adopted to describe the eigenmap distributions of foreground and background. The iterative process will converge when the cost function of the graph-cutting reaches to a stable status [14]. The novelty of the proposed algorithm is the analytic solution of unifying spectral clustering technique with iterative graph cut framework. The efficacy of the proposed image segmentation is evaluated by comparing four sets of manual delineations in 110 ultrasound scans.

2 Spectral Clustering and Iterative Eigenmap Aggregation

The proposed image segmentation method is constituted of two major steps. The first step maps the sonography/ROI into the projected eigenvector subspace as an eigenmap. With the initial spectral clustering result, the second step performs an iterative graph cut framework on the eigenmap by describing the eigenmap distributions of foreground and background with GMMs. The details of each step will be discussed in follows.

2.1 Constrained Normalized Cuts Criterion

The basic idea of graph-based segmentation methods is to seek the best partition from the affinity graph, denoted as G, in which every image pixel is regarded as a graph node and every possible pairwise relation of image pixels represents as a graph edge. The weight of each graph edge is measured by the mutual similarity of the corresponding pixel pair. Denoting that the set of all graph nodes as V, the set of all graph edges as E, and the weighting of all graph edges as W, and image segmentation is simply achieved by partitioning the graph $G = (V, E, W)$ of image into a good set of connected components. Normalized cuts [16] is a graph theoretic criterion for measuring the goodness of an image partition. The measurement of a graph partitioning can be obtained from the set of cutting edges, which represent disjoin node sets. This approach reaches the good partitioning by minimizing the total weight between two different groups, while maximizing the total weight within each of group. The optimization of this approach can be viewed as a solution of the generalized eigenvalue problem that can be efficiently solved in a close form.

To group up similar pixels together based on low-level cues from low to high level structures is so called bottom-up approach. There are two possible situations, which may degrade the performance of graph partitioning. Firstly, object with weak or broken edge may hinder the differentiation process of point sets. Secondly, the process of graph partitioning may also be disturbed by clutter background. On the other hand, those complex background models can be viewed as constraints (or prior knowledge). Yu and Shi [12,13] reformulated normalized cuts criterion as a constrained eigenvalue problem, which integrate both graph partitioning approaches with partial grouping

constraints. There are three types of partial grouping constraints introduced in [12]: 1) unitary generative models, 2) partial grouping, and 3) spatial attention. In this paper, we employed partial grouping constraint, which makes an assumption that the prior expectation of object locations of lesions is located at center position of ROI [17]. Moreover, the similarity between two pixels is defined as the maximum magnitude of intensity edges between them [18].

With the definition of similarity measurement and partial grouping setting, constrained eigenvalue problem based on the normalized cuts criterion can be formally written as:

$$\text{maximize } CNcut(Z) = \frac{1}{K} tr\left(Z^T WZ\right), \tag{1}$$

$$\text{subject to } Z^T DZ = I, \tag{2}$$

$$U^T Z = 0, \tag{3}$$

where $Z = X\left(X^T DX\right)^{-1/2}$ is a scaled partition matrix. X is $N \times K$ partition matrix, that is, $X_{ij} = 1$ if the *i*-th pixel belongs to *j*-th cluster component and 0 otherwise. U forms a set of partial grouping information. The above system can be solved by applying the Rayleigh-Ritz theorem. In order to solve the deficiency caused by sparse cues, Equation (3) can become:

$$U^T PZ = 0, \tag{4}$$

where $P = D^{-1}W$ is the normalized weight matrix. Compare with Equation (3), new partial grouping information includes the property of smoothness and fairness of the local segmentations. Note that it can guide the propagation of biased data points to those neighbors with which it has high similarity. More details about the propagating constraints can be found in [13].

2.2 Iterative Eigenmap Aggregation

The proposed algorithm is composed of two major steps. In first step, we group up similar pixels together with the constrained normalized cut, which is performed to obtain the eigenmap of the constrained eigenvalue system. The summary of first step is as follow. Firstly, the region of interest (e.g. breast lesions) is defined as a rectangle by user from breast sonograms. We then set up the partial grouping constraints around the ROI with 15 pixels width, which make an assumption that the prior expectation of object locations of lesions is located at center position of ROI [17]. Here, affinity matrix is established using the maximum magnitude of intensity edges between two pixels [18]. Finally, the eigenmap is obtained by solving the constrained normalized cuts criterion.

In second step, a higher level framework modeling the weighting distribution of foreground and background is carried out to seek an optimum contour along the lesion. In practice, a breast lesion is delineated by an iterative graph cut framework on the eigenmap. Recently, Rother et al. [14] improved the efficiency of graph cut algorithm [19] and extended graph cut algorithm to color images. Method [14] can be

implemented by following [15], which consists of three steps: initialization, learning GMM parameters, and performing the graph cut framework.

Initialization: According to [20], the eigenmap is automatically scaled into the "hot colormap" ($\mathbf{e}_{\boldsymbol{\lambda}}$). The prior information used in the first step is then marked as known background class ($\alpha = 0$) and all unknown pixels are assigned to foreground class ($\alpha = 1$). By doing so, k Gaussian mixture models (GMMs) is adopted to model the foreground and background, respectively (in this paper, K=5). Note that each component of GMMs includes four elements: 1) the mean (μ); 2) the inverse of the covariance matrix ($\sum^{-1}$); 3) the determinant of the covariance matrix ($\det \sum$); 4) a component weight (π).

Learning GMM parameters: The similarity measurement based on Gaussian components in the foreground or background GMM is calculated for each pixel in the foreground or background class, respectively. Note that each pixel is assigned to the nearest Gaussian component according to the largest similarity. Let component index $\mathbf{K_C}$ be a Gaussian component set which is assigned to each pixel. Hence, the parameters $(\alpha, \mathbf{K_C})$ of GMM for each pixel are created. Accordingly, the new parameters of GMM can also be re-estimated.

Performing the graph cut framework: After learning GMM components, each pixel is assigned to the pair of values $(\alpha, \mathbf{K_C})$. According to [14], the graph cut based energy function can become:

$$E(\alpha, \mathbf{K_C}, \Omega, \mathbf{e}_{\boldsymbol{\lambda}}) = R(\alpha, \mathbf{K_C}, \Omega, \mathbf{e}_{\boldsymbol{\lambda}}) + B(\alpha, \mathbf{e}_{\boldsymbol{\lambda}}), \tag{5}$$

where $\mathbf{e}_{\boldsymbol{\lambda}}$ is the eigenmap via automatic data scaling. Ω indicates as the foreground/background histogram of eigenvector map given α . The smoothness term $B(\alpha, \mathbf{e}_{\boldsymbol{\lambda}})$ can be defined as [15]:

$$B(\alpha, \mathbf{e}_{\boldsymbol{\lambda}}) = 50 \cdot \sum_{\{p,q\} \in N} \frac{e^{-\beta(e_{\lambda p} - e_{\lambda q})^2} \cdot \delta(A_p, A_q)}{dis(p,q)}, \tag{6}$$

and

$$\delta(A_p, A_q) = \begin{cases} 1 & if\ A_p \neq A_q \\ 0 & otherwise \end{cases}. \tag{7}$$

The smoothness term $B_{\{p,q\}}$ is served as a penalty for measuring the discontinuities between two pixels p and q in a specified N-links. The N-link weights are constructed by calculating the similarity between the neighboring pixels in the 8-neighborhood. Given a graph partitioning, the regional term $R(\alpha, \mathbf{K_C}, \Omega, \mathbf{e}_{\boldsymbol{\lambda}})$ is considered as a penalty which is determined by T-links. The foreground T-link weights is constructed

by calculating the similarity between each of pixels to the foreground nodes. Similarly, the background T-link weights is constructed by calculating the similarity between each of pixels to the background nodes. The definition of T-link weights for given a pixel p is formulated as follow [15].

$$R_p\left(\alpha,\mathbf{K_C},\Omega,\mathbf{e_\lambda}\right)=-\log\sum_{i=1}^{|\mathbf{K}|}\left[\pi(\alpha_p,\mathbf{K_C})\frac{1}{\sqrt{\det\Sigma(\alpha_p,\mathbf{K_C})}}\times\right.$$
$$\left.\exp\left(\frac{1}{2}[e_{\lambda p}-\mu(\alpha_p,\mathbf{K_C})]^T\Sigma(\alpha_p,\mathbf{K_C})^{-1}[e_{\lambda p}-\mu(\alpha_p,\mathbf{K_C})]\right)\right], \quad (8)$$

where π are mixture weighting coefficients. The summary of T-links is shown in Table 1.

Table 1. The definition of T-link weights for a pixel p

Pixel Type	Background T-link	Foreground T-link
p∈ source node	0	$\max_p B\left(\alpha_p,e_{\lambda p}\right)$
p∈ sink node	$\max_p B\left(\alpha_p,e_{\lambda p}\right)$	0
p∈ unknown	$R_p\left(Foreground\right)$	$R_p\left(Background\right)$

After building a graph, a standard minimum cut algorithm [21] is applied to seek an optimum segregation of foreground and background nodes. The second stage iterates between *learning GMM parameters* and *performing the graph cut framework* until the classification converges to a stationary status.

3 Results and Discussion

Fig. 1 compares the results of the proposed methods with cell competition algorithm [5,6], level set method [9] and constrained normalized cut [12,13] on a benign lesion.

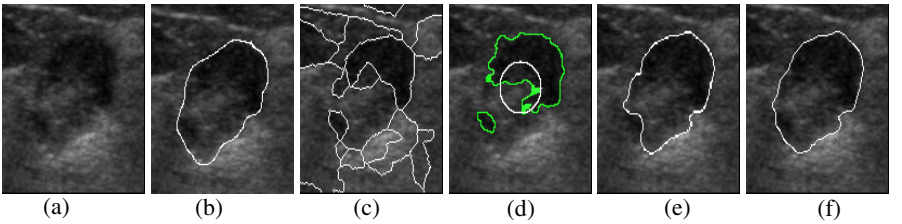


Fig. 1. A comparison of segmentation results from manual drawing, cell competition algorithm, level set method, constrained normalized cut and the proposed algorithm. (a) ROI of a breast sonogram with a benign lesion. (b) A manual delineation. (c-f) Segmentation results of cell competition algorithm, level set method, constrained normalized cut and the proposed algorithm.

It can be observed in Fig. 1(a) that the lower boundary of the benign lesion is relatively difficult to define by comparing to the upper portion. Fig. 1(b) is the manual delineation of this benign lesion by an experienced expert. According to Fig. 1(c) and Fig. 1(d), cell competition algorithm and level set method only can capture partial structure of this lesion. The constrained normalized cut may be able to delineate the major parts of this lesion but still cannot do well in the lower left part, see Fig. 1(e). Distinctively, the proposed algorithm can demarcate this lesion better, as shown in Fig. 1(f). To demonstrate the efficacy of the proposed algorithm, Fig. 2 shows three clips of original breast sonograms. Fig. 2(a) illustrates a malignant breast lesion with shadowing effect and irregular boundary, while the malignant breast lesion in Fig. 2(d) is ill-defined and relatively difficult to perceive. Fig. 2(g) shows a benign breast lesion with edge portions attenuated by acoustic shadows. The derived boundaries of the three cases are depicted in Fig. 2(c), 2(f), and 2(i), respectively. According to Fig. 2, the proposed algorithm seems to be able to effectively identify the breast lesion with complicated shapes. Given the 110 × 110 image/ROI, the proposed algorithm takes around 20 seconds in average to demarcate the boundaries of breast lesions in Intel® Core™2 Quad Processor Q9300.

To evaluate the performance of the proposed algorithm, 110 breast US scans, including 60 benign lesions and 50 malignant lesions, were retrospectively selected as testing data set from the database of Taipei Veterans General Hospital (Taipei, Taiwan). The derived boundaries by the proposed algorithm are compared to four sets of

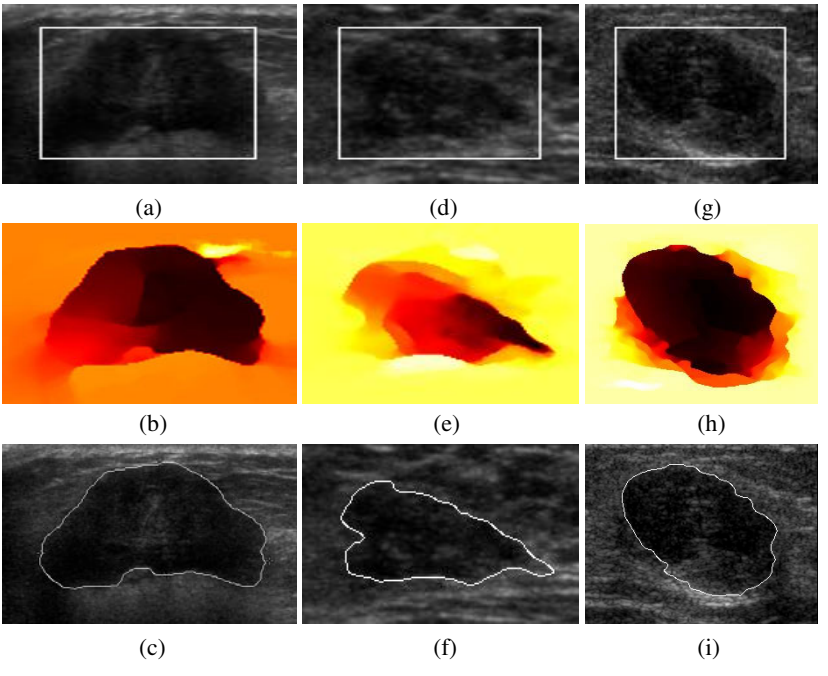


Fig. 2. Illustration of three segmentation results of the proposed algorithm. (a), (d), and (g) are ROIs cropped from three different breast sonograms with partial grouping at image boundaries. (b), (e), and (h) illustrate the projected eigenmap. (c), (f), and (i) are the corresponding segmen-

tation results of (a), (d), and (g), respectively.

Table 2. Performance of the proposed algorithm over 110 breast sonograms by comparing four sets of manual delineations A, B, C, and D. Column COD is mean computer-to-observer distance (COD) and Column IOD is mean maximum interobserver distance (IOD). The COD-IOD suggests if the averaged COD is smaller than IOD. The 95% confidence intervals (CI) of COD-IOD are reported in the fourth column. P is percentages of cases within the interobserver range.

Observer	CO	IO	CO-IO	95% CI	P (%)
A	2.885	4.306	-1.421	(-1.836,-1.006)	87.27%
B	2.969	4.479	-1.510	(-2.062,-0.958)	83.64%
C	4.124	4.613	-0.489	(-0.832,-0.146)	75.45%
D	3.015	4.292	-1.277	(-1.764,-0.789)	86.36%

manual delineations from four experienced experts. The manual delineations were prepared independently without any cross reference. Table 2 reports the comparison with the four sets of manual delineations. It can be found that the variation between algorithm-generated boundaries with manual delineations is less than variation between manual delineations in Table 2. Furthermore, three assessment metrics developed in [22,23] are also involved in the quantitative performance evaluation. The three assessment metrics are Williams Index and overlapping and difference ratios between the algorithm-derived boundaries and the average manual delineations, respectively. The calculated Williams Index is 1.085, which indicates that the algorithm-derived boundaries are relatively stable than the variation between difference experts. The overlapping and difference ratios between the algorithm-derived boundaries and the averaged manual delineations are higher than 0.90 and lower than 0.14, respectively. More details of the three assessment metrics can be found in [22,23]. Accordingly, the boundaries derived by the proposed algorithm deem to be comparable to manual delineations and the proposed algorithm is robust to the problems of weak edges and speckle noise.

4 Conclusions

In this paper, we present a robust segmentation algorithm for the demarcation of lesions in breast sonograms by techniques of spectral eigen-projection and iterative graph-cutting. Firstly, we project the ROI of a breast sonogram onto to a specific eigenmap to reduce the effect of speckle noise and posterior acoustic shadows. The process of project is boosted by the partial grouping setting with the assumption that the object of interest is located at the center of ROI. Following, an iterative graph cut framework is performed to identify the object of interest, i.e. the breast lesion, from the eigenmap. GMMs are adopted to model the gain distributions of the foreground and background on the eigenmap. The proposed method is compared to four sets of manual delineations by four experienced experts. The experiment results corroborates that the derived boundaries of the proposed algorithm are comparable to manual delineations.

References

1. Cheng, J.Z., Chou, Y.H., Huang, C.S., Chang, Y.C., Tiu, C.M., Chen, K.W., Chen, C.M.: Computer-Aided US Diagnosis of Breast Lesions by Using Cell-based Contour Grouping. Radiology 255, 746–754 (2010)
2. Chen, C.M., Chou, Y.H., Han, K.C., Hung, G.S., Tiu, C.M., Chiou, H.J., Chiou, S.Y.: Breast lesions on sonograms: Computer-aided diagnosis with nearly setting-independent features and artificial neural networks. Radiology 226, 504–514 (2003)
3. Stavros, A.T., Thickman, D., Rapp, C.L., Dennis, M.A., Parker, S.H., Sisney, G.A.: Solid Breast Nodules - Use of Sonography to Distinguish Benign and Malignant Lesions. Radiology 196, 123–134 (1995)
4. Noble, J.A.: Ultrasound image segmentation and tissue characterization. Proceedings of the Institution of Mechanical Engineers Part H-Journal of Engineering in Medicine 224, 307–316 (2010)
5. Chen, C.M., Chou, Y.H., Chen, C.S.K., Cheng, J.Z., Ou, Y.F., Yeh, F.C., Chen, K.W.: Cell-competition algorithm: A new segmentation algorithm for multiple objects with irregular boundaries in ultrasound images. Ultrasound in Medicine and Biology 31, 1647–1664 (2005)
6. Cheng, J.Z., Chen, C.M., Chou, Y.H., Chen, C.S.K., Tiu, C.M., Chen, K.W.: Cell-based two-region competition algorithm with a map framework for boundary delineation of a series of 2D ultrasound images. Ultrasound in Medicine and Biology 33, 1640–1650 (2007)
7. Sahiner, B., Chan, H.P., Roubidoux, M.A., Hadjiiski, L.M., Helvie, M.A., Paramagul, C., Bailey, J., Nees, A.V., Blane, C.: Malignant and benign breast masses on 3D US volumetric images: Effect of computer-aided diagnosis on radiologist accuracy. Radiology 242, 716–724 (2007)
8. Chang, R.F., Wu, W.J., Moon, W.K., Chen, D.R.: Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors. Breast Cancer Research and Treatment 89, 179–185 (2005)
9. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Transactions on Image Processing 10, 266–277 (2001)
10. Archip, N., Rohling, R., Cooperberg, P., Tahmasebpour, H.: Ultrasound image segmentation using spectral clustering. Ultrasound in Medicine and Biology 31, 1485–1497 (2005)
11. Noble, J.A., Boukerroui, D.: Ultrasound image segmentation: A survey. IEEE Transactions on Medical Imaging 25, 987–1010 (2006)
12. Yu, S.X., Shi, J.B.: Grouping with Bias. Neural Information Processing Systems (2001)
13. Yu, S.X., Shi, J.B.: Segmentation given partial grouping constraints. IEEE Transactions on Pattern Analysis and Machine Intelligence 26, 173–183 (2004)
14. Rother, C., Kolmogorov, V., Blake, A.: "GrabCut" - Interactive foreground extraction using iterated graph cuts. Acm Transactions on Graphics 23, 309–314 (2004)
15. Talbot, J., Xu, X.: Implementing GrabCut. Brigham Young University (2006)
16. Shi, J.B., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 888–905 (2000)
17. Madabhushi, A., Metaxas, D.N.: Combining low-, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions. IEEE Transactions on Medical Imaging 22, 155–169 (2003)
18. Cour, T., Benezit, F., Shi, J.B.: Spectral segmentation with multiscale graph decomposition. In: CVPR 2005, Washington, DC, USA, pp. 1124–1131 (2005)

19. Boykov, Y., Jolly, M.P.: Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In: International Conference on Computer Vision, vol. I, pp. 105–112 (2001)
20. Imoverlay and imagesc | Steve on Image Processing, http://blogs.mathworks.com/steve/2007/07/20/imoverlay-and-imagesc/
21. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Transactions on Pattern Analysis and Machine Intelligence 26, 1124–1137 (2004)
22. Chalana, V., Linker, D.T., Haynor, D.R., Kim, Y.M.: A multiple active contour model for cardiac boundary detection on echocardiographic sequences. IEEE Transactions on Medical Imaging 15, 290–298 (1996)
23. Chalana, V., Kim, Y.M.: A methodology for evaluation of boundary detection algorithms on medical images. IEEE Transactions on Medical Imaging 16, 642–652 (1997)

Organ Pose Distribution Model and an MAP Framework for Automated Abdominal Multi-organ Localization

Xiaofeng Liu, Marius George Linguraru, Jianhua Yao, and Ronald M. Summers

Imaging Biomarkers and Computer Aided Diagnosis Laboratory, Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, Bethesda, MD 20892

Abstract. Abdominal organ localization is required as an initialization step for most automated abdominal organ analysis tasks, i.e. segmentation, registration, and computer aided-diagnosis. Automated abdominal organ localization is difficult because of the large variability of organ shapes, similar appearances of different organs in images, and organs in close proximity to each other. Previous methods predicted only the organ locations, but not the full organ poses including additionally sizes and orientations. Thus they were often not accurate enough to initialize other image analysis tasks. In this work we proposed a maximum a posteriori (MAP) framework to estimate the poses of multiple abdominal organs from non-contrast CT images. A novel organ pose distribution model is proposed to model the organ poses and limit the search space. Additionally the method uses probabilistic atlases for organ shapes, and Gaussian mixture models for organ intensity profile. An MAP problem is then formulated and solved for organ poses. The method was applied for the localization of liver, left and right kidneys, spleen, and pancreas, and showed promising results, especially on liver and spleen (with mean location and orientation errors under 5.3 mm and 7 degrees respectively).

Keywords: Organ localization**,** maximum a posteriori, probabilistic atlas, pose distribution model.

1 Introduction

The volumes and shapes of abdominal organs can be indicators of disorders, and computed tomography (CT) is commonly adopted for abdominal diagnosis and preoperative planning and guidance. In computer-aided diagnosis (CAD), the identification and segmentation of abdominal organs are essential for further assessment. As in clinical practice, there is a strong CAD incentive for the automated simultaneous detection and analysis of multiple organs, which benefits from inter-organ spatial relationship and interaction, for comprehensive diagnosis.

Many methods have been proposed for the segmentation of individual abdominal organs from contrast-enhanced CT images especially for the liver [1,2,3,4], and recently the simultaneous segmentation of multiple organs [5,6,7]. Most of these methods relied on prior knowledge of the organs, for example probabilistic atlases [8,9], which are sensitive to initialization/registration, and active shape models [1,4]. The knowledge was utilized to provide an initial segmentation and then refined by fitting

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 393–402, 2010.

the shape model and/or using geodesic or parametric deformable models [4]. In addition, multi-dimensional contrast-enhanced CT data were also employed in appearance-based segmentation using component analysis in a Bayesian framework [10], or using a 4D convolution constrained by a historic model of abdominal soft tissue enhancement [6].

Abdominal multi-organ segmentation remains a challenging task because the sizes, shapes and locations of the organs vary significantly in different subjects. Moreover, these organs have similar appearance in CT images, especially non-contrast data, and are in close proximity to each other. Thus the successful segmentation requires a good initial identification and localization of individual organs, generally performed interactively [5,7,8]. Correct organ localization can also benefit other image processing tasks, including registration and computer-aided detection.

Among the most notable automated localization techniques for abdominal organs, Okada et al. [1] initialized the liver segmentation by estimating the abdominal cavity, but it is not certain how well this approach works for smaller organs, e.g., kidneys and pancreas. Yao and Summers [12] used a statistical location model, but the method was limited to estimating only the organ locations without considering the orientations and sizes. Yao et al. [13] simultaneously detected multi-organ locations by finding bounding boxes using principal component analysis and a probabilistic atlas. Due to the large variability of abdominal organ sizes and orientations, however, the location alone cannot completely localize the organs in the abdomen, and thus is not sufficient to accurately initialize other image analysis tasks. Seifert et al. [11] estimated the organ location, orientation, and size using automatically detected anatomical landmarks, semantics and machine learning techniques, but the technical details of the method are not clear.

In this paper, we propose a maximum a posterior (MAP) framework for automated abdominal multi-organ localization. Our method finds the poses of multiple abdominal organs, which include not only the locations, but also the orientations and sizes. The paper first introduces a new abdominal cavity normalization to reduce the variability of organ pose caused by different abdominal sizes across subjects. Next, from a training data set we compute a novel pose distribution model that estimates the probability density functions of organ poses using Parzen windows. The method also uses a probabilistic atlas to model the organ shapes, and a Gaussian mixture model to compute the organ intensity profiles. Last, we formulate an MAP problem and solve it for multi-organ poses from abdominal CT images. The method was applied to 12 data sets (5 organs/set) with promising results.

2 Method

Our method estimates the poses of abdominal organs from non-contrast CT images using a maximum a posteriori (MAP) framework. In this work we focus on five organs, i.e., liver, spleen, left and right kidneys, and pancreas. From a group of N training images $\{\mathbf{I}_n\}_{n=1}^{N}$, these organs are manually segmented and statistically modeled by building a pose distribution model (OPDM), a probabilistic atlas (PA), and a probabilistic intensity profile (IP). For a given subject image, these abdominal organs are then localized using an MAP framework based on these pre-computed statistical models. The method is detailed as follows.

2.1 The MAP Framework

For a given abdominal subject CT image $\mathbf{S}$, the organs are localized by finding the pose parameters that maximize the a posteriori probability, i.e.,

$$\Theta^{(j)*} = \arg\max_{\Theta^{(j)}} p(\Theta^{(j)} \mid \mathbf{S}, O^{(j)}), \tag{1}$$

where $O^{(j)}$ is the j^{th} organ, and $\Theta^{(j)}$ is the pose of $O^{(j)}$ in the given image. The organ pose $\Theta^{(j)}$ is defined using 9 parameters, which include the location $\mathbf{c} = [c_x, c_y, c_z]$, orientation $\mathbf{v} = [v_x, v_y, v_z]$, and scaling $\mathbf{s} = [s_x, s_y, s_z]$. Using Bayes' theorem, the a posteriori probability is re-written as

$$p(\Theta^{(j)} \mid \mathbf{S}, O^{(j)}) = \frac{p(\mathbf{S} \mid \Theta^{(j)}, O^{(j)}) p(\Theta^{(j)} \mid O^{(j)})}{p(\mathbf{S} \mid O^{(j)})}. \tag{2}$$

The denominator in (3) is not related to $\Theta^{(j)}$ and thus can be ignored. The modeling of the prior $p(\Theta^{(j)} \mid O^{(j)})$ and the conditional probability $p(\mathbf{S} \mid \Theta^{(j)}, O^{(j)})$ is explained as follows.

2.2 Abdomen Normalization

To reduce the organ pose variances, the abdomens of different subjects are normalized to account for the shape and size differences. For this, the vertebrae and the ribs are automatically segmented and identified from the CT scans using the method in [12]. Here we consider four vertebrae, T11, T12, L1, and L2, because they span over the location of the majority of the abdominal organs. An abdominal bounding box defines the abdominal cavity and is determined as such: one edge is parallel to the vertebrae line defined by the centers of the four vertebrae and has a length that covers exactly the four vertebrae, and the other edges are defined by finding the minimum bounding box that contains the ribs. The segmented vertebrae and ribs, and the bounding box are illustrated in Fig 1. For normalization, one standard image (denoted as $\mathbf{J}_0$) is chosen from the training images, and its abdominal cavity is considered as the normalized space. All the other images are then normalized to $\mathbf{J}_0$ by aligning the abdominal bounding boxes and are denoted as $\mathbf{J}_n$ for $n = 1, ..., N-1$.

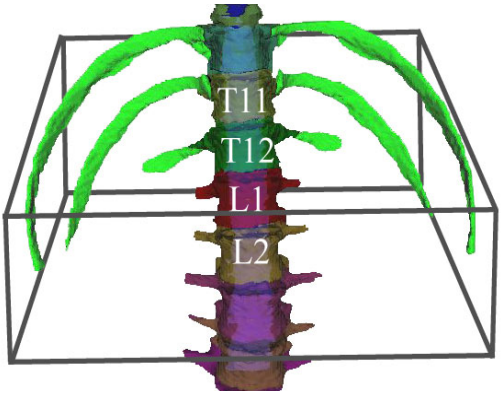


Fig. 1. Illustration of segmented vertebras and ribs, and the abdominal cavity bounding box

2.3 The Prior: Organ Pose Distribution Model

The prior $p(\Theta^{(j)} \mid O^{(j)})$ is modeled using Parzen window [14] technique from the training data sets, and we call it organ pose distribution model (OPDM). The OPDM for each abdominal organ describes the statistical variability of the organ pose, and is built independently.

The pose of organ $O^{(j)}$ in the standard image $\mathbf{J}_0$ is defined as the reference pose, i.e., $\Theta_0^{(j)} = [\mathbf{c}_0^{(j)}, \mathbf{v}_0^{(j)}, \mathbf{s}_0^{(j)}]$ with $\mathbf{c}_0^{(j)}$ being the center of gravity, the organ orientations in the three dimensions being $\mathbf{v}_0^{(j)} = [0,0,0]$, and scales being $\mathbf{s}_0^{(j)} = [1,1,1]$. The organ poses in the other training images relative to the reference pose are computed using a 9-parameter linear registration on the manual segmentations. For each training image $\mathbf{J}_n$ the manually segmented organ is registered to the manual segmentation in $\mathbf{J}_0$ to find the transformation $\mathbf{T}^{(j)}(\cdot)$ with parameters **[u, r, t],** where **u**, **r**, and **t** are the scaling factors, rotation angles, and translation vector respectively. Different organs are registered separately. The organ pose $\theta_n^{(j)}$ in $\mathbf{J}_n$ is then calculated as: $\mathbf{c}_n^{(j)} = \mathbf{T}_n^{(j)}(\mathbf{c}_n^{(j)})$, $\mathbf{v}_n^{(j)} = \mathbf{r}$, and $\mathbf{s}_n^{(j)} = \mathbf{t}$ for $n = 1,...,N-1$.

The 9 pose parameters are assumed independently distributed. Though these parameters may be weakly related, this assumption greatly simplifies the model and reduces the dimensionality of the OPDM. The probability density function of each pose parameter is estimated using Parzen window method. Let θ_k be any of the 9 pose parameters, and $\theta_{k,n}^{(j)}$ be its value in the *n*th image for organ $O^{(j)}$, then

$$p(\theta_k \mid O^{(j)}) = \frac{1}{N} \sum_{n=0}^{N-1} \frac{1}{\sqrt{2\pi}h} e^{-\frac{(\theta_k - \theta_{k,n}^{(j)})^2}{2h^2}}, \tag{3}$$

where *h* is the bandwidth and is estimated using the standard deviation of the data. The OPDM is then constructed as

$$p(\Theta^{(j)} \mid O^{(j)}) = \prod_{k=1}^{9} p(\theta_k \mid O^{(j)}). \tag{4}$$

2.4 Probabilistic Atlas (PA) and Intensity Profile (IP)

To model the conditional probability $p(\mathbf{S} \mid \Theta^{(j)}, O^{(j)})$, a probabilistic atlas (PA) and an intensity profile (IP) are built for each organ. For organ $O^{(j)}$, the PA is constructed using the linear registration results (see Section 2.2) on the manually segmented training images. The probabilistic value on each voxel $\mathbf{x}_i$ in the image volume is computed as the number of training images that $\mathbf{x}_i$ is labeled as part of $O^{(j)}$ divided by the total number of training sets *N*. We denote the PA of O_j as $p(\mathbf{x}_i \mid O^{(j)})$.

The organ intensity profile (IP), denoted as $p(u \mid O^{(j)})$, describes the probability that a voxel in the organ $O^{(j)}$ takes an intensity value of *u*. It is constructed from the

histogram of the Hounsfield unit (HU) value inside the manually segmented organ, and is fitted with a Gaussian mixture model. In the method we use 3 Gaussian components because it produces sufficiently good histogram fitting with low computation.

2.5 The Conditional Probability and Cross Entropy

For simplification, it is assumed that the HU values of voxels inside organs are independent and identically distributed with a known pdf defined as $p(u \mid O^{(j)})$. Thus, the conditional probability in Eqn. (2) can be written as

$$p(\mathbf{S} \mid \Theta^{(j)}, O^{(j)}) = \prod_{\mathbf{x}_i \in \mathbf{V}} p(\mathbf{x}_i \mid \Theta^{(j)}, O^{(j)})\, p(u(\mathbf{x}_i) \mid \Theta^{(j)}, O^{(j)}), \tag{5}$$

where $\mathbf{V}$ is the set of voxels in the image volume, and $u(\mathbf{x}_i)$ is the HU value at $\mathbf{x}_i$ for given pose $\Theta^{(j)}$ ($p(u(\mathbf{x}_i) \mid \Theta^{(j)}, O^{(j)})$ is not dependent on $\Theta^{(j)}$). Let the range of HU values that a voxel can take be $[u_1, u_2, ..., u_M]$. For a given pose $\Theta^{(j)}$, we define the *conditional histogram* $h(u_m \mid \Theta^{(j)}, O^{(j)})$ as the summation of the probability that a voxel belongs to $\Theta^{(j)}$ over all voxels that take an image value of u_m, i.e.,

$$h(u_m \mid \Theta^{(j)}, O^{(j)}) = \sum_{\mathbf{x}_i \in \mathbf{V}} f(u_m, \mathbf{x}_i)\, p(\mathbf{x}_i \mid \Theta^{(j)}, O^{(j)}), \text{ with } f(u_m, \mathbf{x}_i) = \begin{cases} 1, \text{if } u(\mathbf{x}_i) = u_m \\ 0, \text{otherwise} \end{cases}. \tag{6}$$

Thus, Eqn. (5) can be re-written as

$$\begin{aligned} p(\mathbf{S} \mid \Theta^{(j)}, O_j) &= \prod_{m=1}^{M} (p(u_m \mid O^{(j)}))^{h(u_m \mid \Theta^{(j)}, O^{(j)})} \\ &= \exp\left(\sum_{m=1}^{M} h(u_m \mid \Theta^{(j)}, O^{(j)}) \log p(u_m \mid O^{(j)}) \right). \end{aligned} \tag{7}$$

The exponent in Eqn. (7) is the negative *cross entropy* between the two probability functions $h(u_m \mid \Theta^{(j)}, O^{(j)})$ and $p(u_m \mid O^{(j)})$.

2.6 Pose Estimation Using the MAP

The MAP problem in (1) and (2) can be simplified using Eqns. (4) and (7). As the denominator in Eqn. (2) is not related to $\Theta^{(j)}$, it can be ignored from the MAP estimation. In addition, the maximization of Eqn. (1) is equivalent to maximizing the logarithm of the a posteriori probability (see Eqn. (2)). Thus the MAP problem in Eqn. (1) is equivalent to solving

$$\Theta^{(j)*} = \arg\max_{\Theta^{(j)}} \left\{ -H(h(u_m \mid \Theta^{(j)}, O^{(j)}),\, p(u_m \mid \Theta^{(j)}, O^{(j)})) + \sum_{k=1}^{9} \log p(\theta_k \mid O^{(j)}) \right\}, \tag{8}$$

with $H(\cdot, \cdot)$ being the cross entropy.

Eqn. (8) can be solved for the organ pose $\Theta^{(j)*}$ using any gradient-based optimization method. Here we used the steepest descent method [15]. Different organs

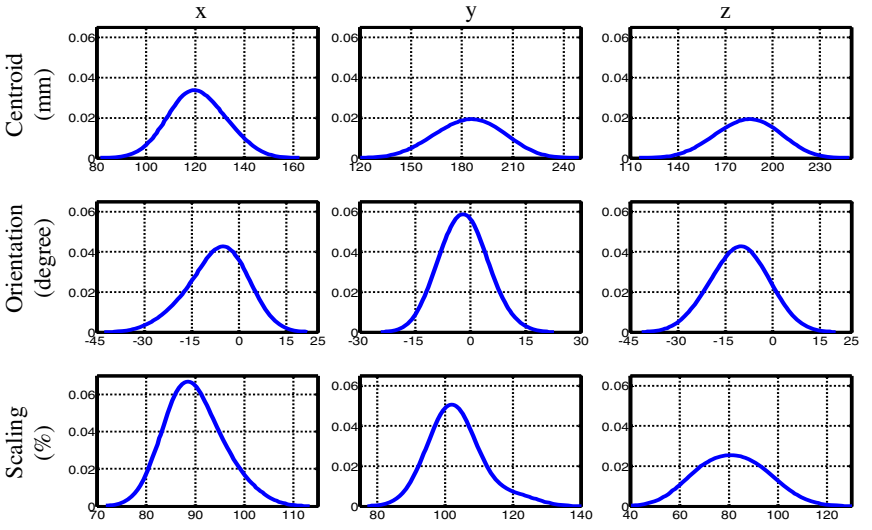


Fig. 2. The pose distribution functions of liver. From top to bottom are the centroid, orientation, and scaling respectively; the *x*, *y*, and *z* components are shown from left to right.

are localized independently. A multi-resolution strategy is adopted to reduce the computation. In addition, the organ center, orientation, and scaling factor are optimized sequentially for faster convergence and to avoid local optima.

3 Results

The method was applied to abdominal non-contrast CT scans of 12 (6 males and 6 females) patients. Data were collected with a LightSpeed Ultra scanner (GE Healthcare). The slice thickness was 1 mm and the in-slice resolution varied from 0.54 mm to 0.77 mm. In all the 12 images, the five organs (liver, left and right kidneys, spleen, and pancreas) were manually segmented by a medical student and supervised by a radiologist. For validation, 12 experiments were performed with a leave-one-out strategy. In each experiment, one dataset was picked as the subject image, and the remaining 11 datasets served as the training data and were used to build the OPDM and PA. After the organ pose was estimated, the probability atlas was transformed to the subject image using the estimated pose. For quantitative evaluation, the pose errors were calculated by comparing the localized poses with true poses, which were computed from the linear registration as described in Sec. 2.2. The Dice's coefficient [16] was used to measure the symmetric volume overlap between the estimated organs and manual segmentations.

Fig. 2 shows the pose distribution functions, exemplified for the liver, computed using all 12 data sets. Note the large variance of the distributions especially for the y- and z-positions and z-scaling. Fig. exemplifies the organ localization results on one dataset using selected axial slices. **Error! Reference source not found.** presents the organ pose errors between the organ localization results and the true poses computed

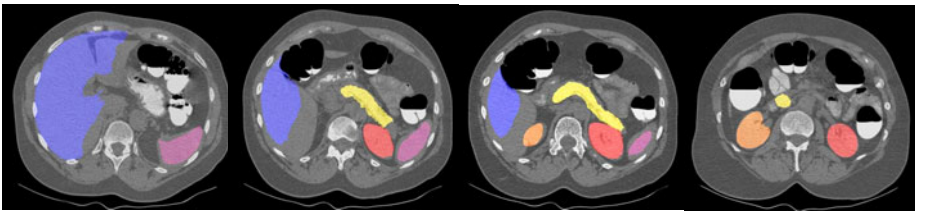


Fig. 3. The localization results of five organs on one data set. Images are axial slices from the image volume with the organs identified using different colors.

Table 1. The pose estimation errors between the automated localization results and the true poses. In brackets we present the kidney pose errors without considering the outlier cases.

Organ \ Pose		Centroid (mm)			Orientation (degree)			Scale (%)		
		x	y	z	x	y	z	x	y	z
Liver	mean	2.42	2.23	4.61	4.09	2.16	3.50	4.19	5.94	18.8
	std	2.38	1.66	2.86	3.09	1.89	2.61	3.70	4.98	6.83
Spleen	mean	5.28	4.62	3.70	6.89	5.94	4.93	8.31	18.8	24.2
	std	5.63	4.11	3.35	6.05	4.56	3.72	6.60	8.85	17.2
Left Kidney	mean	1.06 (0.73)	1.78 (1.34)	9.63 (2.17)	3.77 (3.98)	4.01 (2.91)	3.90 (3.48)	6.81 (6.27)	7.18 (7.62)	4.29 (4.72)
	std	1.00 (0.56)	1.50 (0.98)	17.7 (1.73)	4.35 (4.74)	3.72 (2.92)	4.45 (4.20)	4.41 (4.66)	5.80 (6.13)	3.27 (3.31)
Right Kidney	mean	1.31 (1.08)	2.69 (1.99)	8.69 (6.40)	3.13 (2.40)	4.42 (4.20)	3.45 3.63)	7.35 (7.00)	10.04 (9.93)	4.58 (4.41)
	std	1.31 (1.07)	3.24 (2.23)	10.47 (7.16)	3.26 (2.17)	3.88 (3.99)	3.92 (4.06)	3.61 (3.58)	7.69 (8.06)	3.84 (3.99)
Pancreas	mean	5.64	6.39	4.17	6.23	5.05	6.09	9.20	10.89	7.79
	std	4.78	7.38	3.68	6.74	4.62	4.41	7.45	12.71	7.64

according to Section 2.2. Over all, the method performed well for the liver, spleen, left and right kidneys, but less adequate for the pancreas, an organ with higher variability across subjects. In particular for the liver and spleen, the method worked consistently well on all the 12 experiments with small variations. There is a visible correlation between the magnitude of the liver pose errors and the variance of the distributions in Fig. 2. For the left kidney, there were two cases where the kidneys partly overlapped with the nearby spleen (outliers). This was caused by the very similar HU value profiles of spleen and kidneys in non-contrast CT, and their spatial proximity. Left kidneys were well localized in the other 10 cases. For the right kidney, there was one outlier where the method partly converged to the liver for similar reasons. These outlier cases suffered from high location errors in the z-direction of the centroid estimation.

Table 1 also shows that the scale estimation has a larger error than the centroid and orientation especially for the liver and the spleen. This is mainly because the scales of these organs present a large variability in the training dataset, as exemplified for the liver scale in z-axis for liver (Fig. 2). We expect these errors to be reduced with a larger training data set.

Table 2. The symmetric volume overlap (Dice's coefficient) after organ localization

	Liver	Spleen	Left Kidney	Right Kidney	Pancreas
mean	0.77	0.69	0.73	0.74	0.43
std	0.03	0.06	0.21	0.14	0.15

Shows the Dice's coefficient results for the five organs from the leave-one-out experiments. Dice's coefficient is a common measure of segmentation accuracy with a value of 1.0 meaning that the localized organ completely overlaps with the manual segmentation. Please note that our method localizes abdominal organs and does not address segmentation, hence the inherent inter-patient organ shape variability affects the percentage of overlap. For the left kidney, the two outlier resulted in very low Dice values (0.21 and 0.33), while the other 10 cases had a Dice's coefficient of 0.82±0.03. The pancreas had a low Dice value in general because of its thin and variable shape. Moreover, the manual segmentation of pancreases was less reliable as it is often difficult to differentiate the pancreas from surrounding tissues in CT data without contrast.

Our method was implemented in C++ and run on a computer with 2GB RAM and 3GHz CPU. In our implementation, the localization of liver took about 1 minute, and the localization of each of the other four organs took about 20 seconds. The liver localization takes longer because it has a larger size and thus requires more computation (see Eqn. (6)).

4 Discussion

The proposed organ localization method employed a pose distribution model to limit the searching space, and an MAP framework to estimates the locations, orientations and scales of abdominal organs. The organ localization can be used as an initialization of other medical image analysis tasks in abdominal CAD, e.g., segmentation and registration. The results showed a reasonably good overlap with manual segmentations on liver, spleen, and kidneys, and suggested it can work as an accurate initialization of segmentation or registration tasks.

Comparing with other organ localization methods [1,12,13], the proposed method estimates the organ orientations and scales in addition to the locations and thus produces more accurate results. For example, although Yao and Summers' method [12] estimated organ locations with similar accuracy as the proposed method, their method resulted a much poorer organ localization results because of the lack of knowledge in organ orientations and scales (by visually checking the volume overlaps from Fig. 4 in [12]).

The method performed consistently well on the liver and the spleen because they have larger sizes, which helps the MAP optimization to converge correctly. The failed cases of kidneys were mainly due to their smaller sizes and proximity to organs with similar intensity profiles. In addition, the localization of pancreas was more challenging because of its irregular shape and large pose variance across subjects. In this work, different organs were localized independently. In future work, we will further

improve the method through modeling inter-organ relations and interactions so that smaller organ localizations can benefit from the localization results of their neighbors.

In the future we will use more data to build the models and expect to achieve a more precise pose distribution function and more accurate results. In addition, in non-contrast CT images, the HU values of abdominal soft tissues are similar, which makes it difficult to correct the localization and differentiate organs based on HU values. Better localization results are expected when applying the method to contrast-enhanced CT images, and it is our future work. In addition, we will incorporate the proposed method with atlas-based segmentation methods for fully automatic organ segmentation.

In conclusion, we developed a novel abdominal multi-organ localization method from non-contrast CT images using an MAP framework. The method computes the organ pose distributions, organ probabilistic atlases, and organ intensity profiles from training data after the abdominal cavity is normalized. It then formulates an MAP problem and solves it for organ poses in given CT images. The method was applied to five organs (liver, left and right kidneys, spleen, and pancreas) with promising results.

Acknowledgements

This work was supported by the Intramural Research Program of the National Institutes of Health, Clinical Center. The authors thank Jesse K. Sandberg and Javed Aman for helping with the data analysis.

References

1. Okada, T., et al.: Automatic segmentation of the liver from 3D CT images using probabilistic atlas and multilevel statistical shape model. Academic Radiology 15, 1390–1403 (2008)
2. Song, Y., et al.: Liver segmentation using automatically defined patient specific B-spline surface models. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part II, LNCS, vol. 5762, pp. 43–50. Springer, Heidelberg (2009)
3. Linguraru, M.G., et al.: Atlas-based automated segmentation of spleen and liver using adaptive enhancement estimation. Med. Phys. 37(2), 771–783 (2010)
4. Heimann, T., Meinzer, H.-P.: Active shape models for a fully automated 3D segmentation of the liver – an evaluation on clinical data. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 41–48. Springer, Heidelberg (2006)
5. Shimizu, A., et al.: Segmentation of multiple organs in non-contrast 3D abdominal CT images. Int. J. Comp. Assist. Radiol. Surg. 2, 135–142 (2007)
6. Linguraru, M.G., Summers, R.M.: Multi-organ segmentation in 4D contrast-enhanced abdominal CT. In: IEEE Int. Symp. Biomed. Imaging (ISBI), pp. 45–48 (2008)
7. Okada, T., et al.: Construction of hierarchical multi-organ statistical atlases and their application to multi-organ segmentation from CT images. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 502–509. Springer, Heidelberg (2008)
8. Park, H., Bland, P.H., Meyer, C.R.: Construction of an abdominal probabilistic atlas and its application in segmentation. IEEE Trans. Med. Imaging 22(4), 483–492 (2003)

9. Reyes, M., et al.: Anatomical Variability of Organs via Principal Factor Analysis from the Construction of an Abdominal Probabilistic Atlas. In: IEEE International Symposium on Biomedical Imaging (ISBI), pp. 682–685 (2009)
10. Hu, X., et al.: Independent analysis of four-phase abdominal CT images. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3217, pp. 916–924. Springer, Heidelberg (2004)
11. Seifert, S., et al.: Hierarchical parsing and semantic navigation of full body CT data. In: Proc. SPIE, vol. 7259, pp. 725902-725902-8 (2009)
12. Yao, J., Summers, R.M.: Statistical location model for abdominal organ localization. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part II. LNCS, vol. 5762, pp. 9–17. Springer, Heidelberg (2009)
13. Yao, C., et al.: Simultaneous location detection of multi-organ by atlas-guided eigen-organ method in volumetric medical images. Int. J. Comp. Assist. Radiol. Surg. 1, 42–45 (2006)
14. Parzen, E.: On the estimation of a probability density function and the more. Annals of Math. Stat. 33, 1065–1072 (1962)
15. Burden, R.L., Faires, J.D.: Numerical analysis. Brooks Cole, Pacific Grove (2004)
16. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology 26(3), 297–302 (1945)

Probabilistic Refinement of Model-Based Segmentation: Application to Radiation Therapy Planning of the Head and Neck

Arish A. Qazi[1], John J. Kim[1,2], David A. Jaffray[1,3], and Vladimir Pekar[4]

[1] Radiation Medicine Program, Princess Margaret Hospital, Toronto, ON, Canada
[2] Dept. of Radiation Oncology, University of Toronto, Toronto, ON, Canada
[3] Dept. of Medical Biophysics, University of Toronto, Toronto, ON, Canada
[4] Philips Research North America, Markham, ON, Canada
arish.qazi@rmp.uhn.on.ca

Abstract. Radiation therapy planning requires accurate delineation of target volumes and organs at risk. Traditional manual delineation is tedious, and can require hours of clinician's time. The majority of the published automated methods belong to model-based, atlas-based or hybrid segmentation approaches. One substantial limitation of model-based segmentation is that its accuracy may be restricted either by the uncertainties in image content or by the intrinsic properties of the model itself, such as prior shape constraints. In this paper, we propose a novel approach aimed at probabilistic refinement of segmentations obtained using 3D deformable models. The method is applied as the last step of a fully automated segmentation framework consisting of automatic initialization of the models in the patient image and their adaptation to the anatomical structures of interest. Performance of the method is compared to the conventional model-based scheme by segmentation of three important organs at risk in the head and neck region: mandible, brainstem, and parotid glands. The resulting segmentations are validated by comparing them to manual expert delineations. We demonstrate that the proposed refinement method leads to a significant improvement of segmentation accuracy, resulting in up to 13% overlap increase.

1 Introduction

With recent advances in Intensity Modulated Radiation Therapy (IMRT), it is possible to deliver a precise dose to the tumor while minimizing the damage to the surrounding healthy tissue [1]. This requires accurate contouring of all the structures to be treated as well as the organs that are at risk of receiving harmful radiation on planning computed tomography (CT) images. Conventionally applied manual slice-wise contouring is often tedious and time consuming. Treatment planning in the head and neck area is especially complex and can require hours of contouring work.

Auto-segmentation methods in radiation therapy planning (RTP) mostly fall into the categories of i) deformable or model-based approaches, ii) atlas-based

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 403–410, 2010.

methods, and iii) hybrid approaches. Methods based on classical deformable models [2,3] rely on local image features, such as edges and make use of the assumption that object boundaries are distinct. In the areas where this assumption does not hold, e.g. soft tissue in CT images, these methods may fail. Prior information about shape, gray appearance, etc. can be included to improve robustness [4,5]; however, finding an appropriate balance between the influence of image content and prior information is generally not easy and may limit the segmentation accuracy. Model initialization is typically done manually by defining seed points or by a drag-and-drop operation.

Recently, atlas-based auto-segmentation methods have gained popularity in RTP [6,7]. Such methods rely on one or several atlas images, which contain contours labeled by an expert. To segment a new clinical case, the atlas is registered to the image and the structures of interest are then transformed using the mapping determined by the registration. The advantage of this approach lies in its universality which allows handling of a wide variety of segmentation problems. On the other hand, the resulting segmentation is directly linked to the accuracy of image registration, which is often very challenging due to high variability of the patient anatomy, organ motion, and image artifacts.

Hybrid approaches combine registration and segmentation into a common framework [8,9], where, for example, evolution of deformable models can serve as a registration constraint or used to compensate for the residual differences after the registration step.

In this paper, we apply a hybrid approach which uses a combination of point-based registration and model-based segmentation (MBS), and introduce a novel refinement step to improve the results of MBS. It assumes that an uncertainty area exists around the boundary defined by the deformable model and that subsequent classification of this region into object / non-object classes will improve the segmentation accuracy. This concept is implemented by the following steps. First, a probabilistic mask is created by averaging the registered expert segmentations. This mask is next registered with the result of MBS in a new patient, and the uncertainty area is refined using voxel classification based on a plurality of low-level image features. The proposed refinement approach is generic and can be applied in combination with other model-based segmentation methods.

2 Method

2.1 Probabilistic Mask Construction

Head and neck CT images of 25 patients were acquired using a standard field of view, and did not contain large neck deformations due to disease (N0 necks). Scan resolution for all datasets was approximately $1\times1\times2$ mm^3. For probabilistic mask construction and voxel classification training, 15 datasets were used, and the remaining 10 datasets were used to validate the method.

In order to construct the probabilistic masks, manual expert delineations of three important organs at risk in the head and neck region: mandible, brainstem, and parotid glands were used. The expert delineations were available in

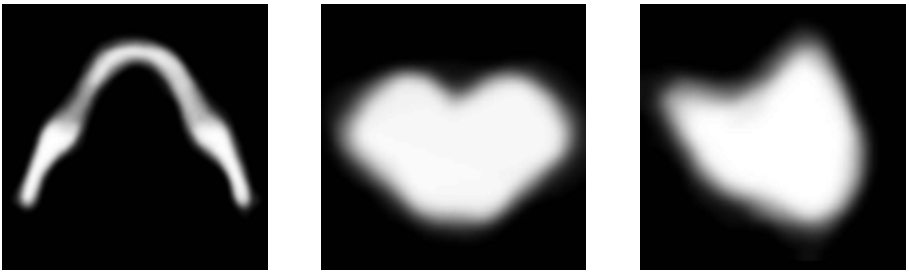


Fig. 1. Axial slices of the probabilistic masks for the mandible (left), brainstem (middle) and left parotid (right). Brighter areas correspond to higher probabilities.

the form of binary volumes and were subsequently converted to triangular surface meshes to facilitate the registration process. The first step in mask construction is point-based affine registration of all surface meshes with an unbiased mean mesh. Analogously to [6], the unbiased mesh is constructed by registration of all meshes to an arbitrary reference mesh to create an initial mean mesh and then iterating the process each time taking the resulting mean mesh from each iteration as the reference mesh for registration. On convergence, we apply the resultant transformations to align the binary volumes and average the results to construct a probabilistic mask. The areas of intersection of all binary volumes are assigned the probability of 1, and the regions with the probabilities between 0 and 1 represent the areas of uncertainty w.r.t. affine transformations, see Fig. 1. The probabilistic mask is blurred with the Gaussian kernel ($\sigma = 1\,\text{mm}$), which has the effect of simulating organ variability and smoothing the borders.

2.2 Model-Based Segmentation and Probabilistic Mask Registration

The implemented model-based segmentation approach uses the energy minimization technique adapted from [4]. The organ models, used for adaptation, are created from manual expert segmentations of a single reference dataset and are represented as triangular surface meshes with integrated prior information from both shape and appearance. In contrast to the original approach, where model placement in the image to segment was performed manually by drag-and-drop, fully automatic model initialization is used in this work. It is based on the registration method developed in [10] for automatic retrieval of corresponding anatomical point landmarks to generate a thin-plate spline transformation between the reference and target image. To retrieve the landmark correspondences, an optimization procedure is applied, aimed at minimizing the gray value difference between image patches around a set of 46 manually selected point landmarks covering areas of head and neck anatomy with characteristic appearance, e.g. anterior clinoid processes, mastoid processes, and cervical vertebra tips.

The adaptation of the reference models to the unseen dataset is carried out by minimizing the sum of external energy attracting the model to the image

Table 1. Candidate feature list for voxel classification; $H(i)$ denotes the local histogram and L is the number of gray levels

Feature	Formulation
Voxel intensity and position	$I(x, y, z),\ x, y, z$
Moments of intensity histogram	$\mu = \sum_{i=0}^{L-1} iH(i), \sqrt{\sum_{i=0}^{L-1}(i-\mu)^2 H(i)}$, $\sum_{i=0}^{L-1}(i-\mu)^3 H(i), \sum_{i=0}^{L-1}(i-\mu)^4 H(i)$
Entropy and uniformity	$-\sum_{i=0}^{L-1} H(i) \log H(i), \sum_{i=0}^{L-1} H(i)^2$
Intensity smoothed	$I_\sigma(x, y, z)$
1st and 2nd order derivatives	$I_x, I_y, I_z, I_{xx}, I_{yy}, I_{zz}, I_{xy}, I_{xz}, I_{yz}$
Magnitude of Hessian eigenvalues	$\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}$
Curvature and Laplacian	$\lambda_1\lambda_2\lambda_3, \lambda_1 + \lambda_2 + \lambda_3$
Gradient magnitude	$\sqrt{I_x^2 + I_y^2 + I_z^2}$

edges and internal energy maintaining the consistency of model shape. After adaptation, the mean meshes of the three organs, constructed as described in Section 2.1 are registered with the adapted reference meshes using affine transformations. The resulting transformations are then applied to the probabilistic masks to transfer them to the image.

2.3 Refinement by Voxel Classification

In the refinement step, voxels in the transformed probabilistic masks, with the probabilities between 0 and 1, representing the uncertainty around the segmented boundary are classified into organ and background class using local low-level image features. For classification, a fast implementation of a kNN classifier based on approximate nearest neighbor search was used [11], with k chosen to be 25.

Features. The performance of the classifier is highly dependent on how well the features are able to discriminate a certain tissue type from the surrounding structures. Image features used in our method are discussed below.

Raw image intensity and 3-D position: The motivation for including these features is that when manually segmenting an organ of interest, a clinician always looks for the location and the intensity of the organ in the image.

Multi-scale image intensities and derivatives: Three scales (1.7 mm, 2.0 mm, and 2.25 mm) are chosen to cover the variation in organ size. Each scale represents the standard deviation of the Gaussian kernel.

Local texture and structure properties: These features are based on the gray-level distribution, as quantified from the local intensity histogram. The features represent the first few moments of the histogram, such as mean, standard deviation, skewness, and kurtosis. Additionally, local entropy and uniformity

are also included. The features are quantified for each voxel locally, using a neighborhood size of 5×5×3 pixels (approx. 5×5×6 mm), chosen experimentally.

Shape based features: The magnitude of the Hessian eigenvalues as a measure of object contrast is included. Rotationally invariant features, such as Laplacian, Gaussian curvature, and gradient magnitude are also included.
This results in an initial set of 52 candidate features at three scales, see Table 1. All features are normalized to a zero mean and unit variance to ensure that the classifier is not sensitive to scaling.

Feature Selection. A high-dimensional feature space not only increases the computational time but can also degrade the classification performance. We employ a feature selection step based on sequential forward floating selection (SFFS) [12] to choose the optimal sub-set of features. This translates to a forward selection (FS) step followed by backward selection (BS). FS starts from an empty set and adds features sequentially as long as the performance criterion improves. Subsequently, BS iteratively removes the least significant features according to the performance criterion. The outcome of the performance criterion is evaluated at each iteration and we stop iterating when the dimensionality of the feature space reaches a point after which the improvement is not significant. For the performance criterion, we maximize the area under the receiver operating characteristic (ROC) curve. The ROC curve is determined by varying the threshold for the classifier and then plotting the ratio of false positives vs. the ratio of true positives [13]. The feature selection is carried out by randomly dividing the training data into two sub-sets. The classifier is trained for a certain combination of features on the first set and the performance is then evaluated on the second set. After the feature selection, the classifier is constructed from all the training data, using the optimally selected features only. To avoid redundant data and to speed up the computation, 60% of randomly selected voxels were sampled from each expert segmentation and background to train the classifier.

Selected features: The Gaussian smoothed intensities at all scales, along with the raw intensity and the position in the image were selected for all three organs. Features selected for the brainstem and the parotids were quite similar, and primarily consisted of second order derivatives, Gaussian curvature and Laplacian. Features selected for the mandible, however, consisted of local textural features, first order derivatives, and gradient magnitude.

Estimation of Posterior Probability and Classification. For each voxel v to be classified, we compute the feature vector $\mathbf{F}_v$. The posterior probability of v belonging to class w_i is then computed by $p(w_i|\mathbf{F}_v) = k_i/k_T$, where k_T is the total number of neighbors, and k_i, determined by the classifier, is the number of nearest neighbors belonging to class w_i. Computation of posterior probabilities leads to a "soft" classification of the voxels. In order to obtain a hard segmentation we define a cutoff point, chosen to be 0.5. The prior probability from the mask is included into the computation of the final probability to enforce

Table 2. Segmentation results (in terms of DSC), comparing MBS and the refinement approach vs. manual expert segmentations. Mean percentage DSC improvement (last row) achieved through refinement.

Dataset	Mandible (MBS)	Mandible (Refined)	Brain-stem (MBS)	Brain-stem (Refined)	Left Parotid (MBS)	Left Parotid (Refined)	Right Parotid (MBS)	Right Parotid (Refined)
1	0.876	0.914	0.832	0.867	0.643	0.698	0.672	0.747
2	0.867	0.916	0.816	0.860	0.696	0.710	0.597	0.663
3	0.864	0.867	0.876	0.898	0.692	0.753	0.658	0.755
4	0.886	0.893	0.861	0.885	0.754	0.828	0.733	0.829
5	0.873	0.905	0.808	0.820	0.696	0.781	0.700	0.720
6	0.860	0.889	0.848	0.875	0.739	0.795	0.754	0.808
7	0.883	0.910	0.866	0.882	0.691	0.745	0.656	0.727
8	0.888	0.913	0.808	0.856	0.714	0.747	0.667	0.699
9	0.877	0.901	0.851	0.873	0.826	0.866	0.749	0.841
10	0.864	0.911	0.844	0.876	0.723	0.804	0.701	0.805
ΔDSC(%)		+4.9		+4.0		+9.9		+13.0

a shape prior, which helps to minimize noise and results in a relatively smooth segmentation. Thus our decision rule is defined as:

$$v \in \begin{cases} w_o, & l_1 p(w_o|\mathbf{F}_v) + l_2 p(m) > 0.5 \\ w_b, & \text{otherwise,} \end{cases}$$

where w_o is the organ class, and w_b is background, $p(m)$ is the probability from the mask, and l_1, l_2 are the weighting factors, both chosen to be 0.5.

3 Results

To quantitatively evaluate the resultant segmentations, we compared them to the manual segmentations. Additionally, we also compared the results to the non-refined MBS approach. The evaluation was carried out by estimating the volume overlap fraction known as the dice similarity coefficient (DSC) on 10 testing datasets. Table 2 summarizes the results: for each structure, the first column lists the DSC for MBS, followed by the DSC of our refinement method in the second column. Fig. 2 shows the segmentation results on a sample dataset, obtained both by MBS and the refinement method. Note the inaccuracies in MBS due to the built-in shape constraints which do not allow the model to precisely follow even sharply defined edges in the image, see e.g. the third row of Fig. 2.

In order to give an estimate of computation time a distinction is made between "offline" and "online" calculations. The offline calculations, consisting of probabilistic mask construction and feature selection are performed only once. Online calculations are carried out for each new patient and include the automatic initialization of the models, their adaptation, and mask refinement. The entire online calculation for the 3 organs, took about 3 min. on a 2.8 GHz dual-core AMD processor, in which the classification step took approximately 30 sec.

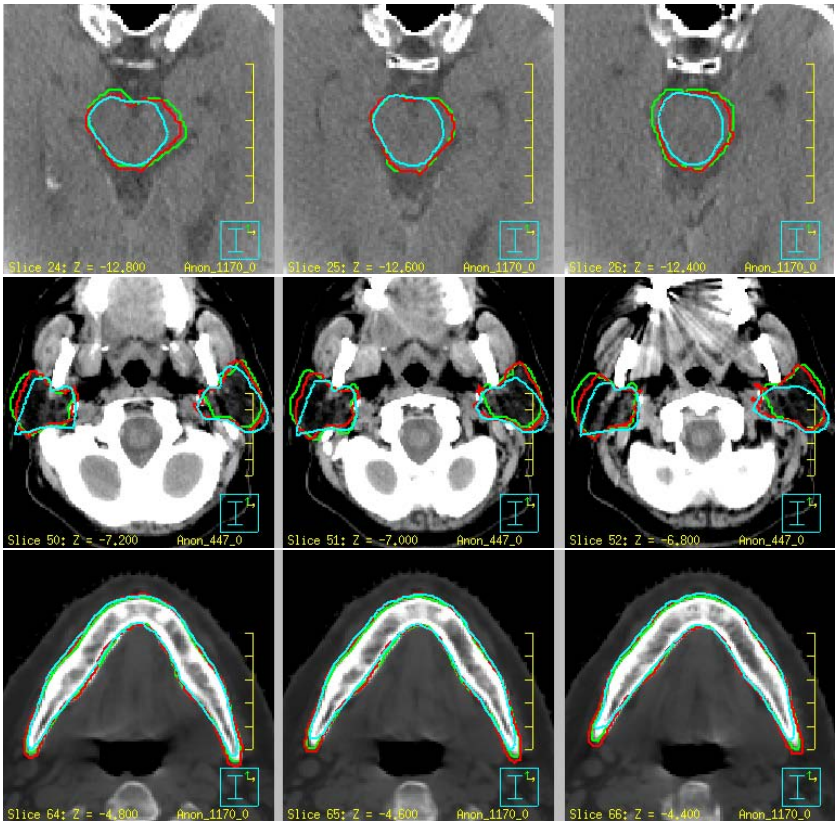


Fig. 2. From top to bottom: exemplary results of probabilistic refinement (red) compared to model-based segmentation (light blue) and expert delineations (green) for brainstem, parotid glands, and mandible

4 Discussion

We have introduced a novel approach to refinement of the existing model-based segmentation. It is implemented by classification of the uncertainty area, quantified from a probabilistic atlas, using a complex voxel labeling step. The method incorporates a feature selection step which automatically extracts the optimal set of features from a pool of textural and geometrical features at different scales.

The approach was quantitatively validated to segment three important organs at risk in the context of radiation therapy planning of the head and neck: mandible, brainstem and parotid glands. For all structures tested, the refined method clearly outperforms the standalone MBS approach, both quantitatively and by visual appreciation of the results. The lower DSC for the parotid glands compared to other tested structures may be attributed to their high shape variability and may be improved by increasing the size of the training data.

Our method is generic and can potentially be applied to any application involving segmentation based on deformable models. Combining the advantages of both local low-level features and global high-level prior shape information is a first step towards achieving a more reliable and accurate segmentation. Future work will involve evaluating the method on other anatomical structures, and evaluating its sensitivity to different registration techniques for creation of the probabilistic masks. We also plan to implement a multi-atlas segmentation scheme, where multiple models derived from different patients can be used to better conform to the anatomy of a particular patient.

References

1. Chou, W.W., Puri, D.R., Lee, N.Y.: Intensity-modulated radiation therapy for head and neck cancer. Expert Rev. Anticancer Ther. 5, 515–521 (2005)
2. Honea, D.M., Snyder, W.E.: Three-dimensional active surface approach to lymph node segmentation. Proc. SPIE Med. Imag. 3661, 1003–1011 (1999)
3. Yan, J., Zhuang, T., Zhao, B., Schwartz, L.H.: Lymph node segmentation from CT images using fast marching method. Comput. Med. Imag. Graph. 28, 33–38 (2004)
4. Pekar, V., McNutt, T.R., Kaus, M.R.: Automated model-based organ delineation for radiotherapy planning in prostatic region. Int. J. Radiat. Oncol., Biol., Phys. 60, 973–980 (2004)
5. Freedman, D., Radke, R.J., Zhang, T., Jeong, Y., Lovelock, D.M., Chen, G.T.Y.: Model-based segmentation of medical imagery by matching distributions. IEEE Trans. Med. Imag. 24, 281–292 (2005)
6. Commowick, O., Gregoire, V., Malandain, G.: Atlas-based delineation of lymph node levels in head and neck computed tomography images. Radioth. Oncol. 87, 281–289 (2008)
7. Han, X., Hoogeman, M.S., Levendag, P.C., Hibbard, L.S., Teguh, D.N., Voet, P., Cowen, A.C., Wolf, T.K.: Atlas-based auto-segmentation of head and neck CT images. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 434–441. Springer, Heidelberg (2008)
8. Gorthi, S., Duay, V., Houhou, N., Bach Cuadra, M., Schick, U., Becker, M., Allal, A.S., Thiran, J.P.: Segmentation of head and neck lymph node regions for radiotherapy planning using active contour-based atlas registration. IEEE J. on Sel. Top. Sig. Proc. 3, 135–147 (2009)
9. Zhang, X., Tian, J., Wu, Y., Zheng, J., Deng, K.: Segmentation of head and neck CT scans using atlas-based level set method. MIDAS J. (2009), http://www.midasjournal.org/browse/publication/668
10. Leavens, C., Vik, T., Schulz, H., Allaire, S., Kim, J., Dawson, L., O'Sullivan, B., Breen, S., Jaffray, D., Pekar, V.: Validation of automatic landmark identification for atlas-based segmentation for radiation treatment planning of the head-and-neck region. In: Proc: SPIE Med. Imag., vol. 6914 (2008)
11. Arya, S., Mount, D.M., Netanyahu, N.S., Silverman, R., Wu, A.Y.: An optimal algorithm for approximate nearest neighbor searching fixed dimensions. J. ACM 45, 891–923 (1998)
12. Pudil, P., Novoviova, J., Kittler, J.: Floating search methods in feature selection. Pat. Recog. Let. 15, 1119–1125 (1994)
13. Zweig, M.H., Campbell, G.: Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin. Chem. 39, 561–577 (1993)

Skin Lesions Classification with Optical Spectroscopy

Asad Safi, Victor Castaneda, Tobias Lasser, and Nassir Navab

Computer Aided Medical Procedures (CAMP),
Technische Universität München, Germany
{safi,castaned,lasser}@in.tum.de, navab@cs.tum.edu

Abstract. Diagnosis of benign and malign skin lesions is currently mostly relying on visual assessment and frequent biopsies performed by dermatologists. As the timely and correct diagnosis of these skin lesions is one of the most important factors in the therapeutic outcome, leveraging new technologies to assist the dermatologist seems natural. Complicating matters is a blood cancer called Cutaneous T-Cell Lymphoma, which also exhibits symptoms as skin lesions. We propose a framework using optical spectroscopy and a multi-spectral classification scheme using support vector machines to assist dermatologists in diagnosis of normal, benign and malign skin lesions. As a first step we show successful classification (94.9%) of skin lesions from regular skin in 48 patients based on 436 measurements. This forms the basis for future automated classification of different skin lesions in diseased patients.

Keywords: Skin cancer, Optical spectroscopy, Classification.

1 Introduction

Skin cancer is one of the most common cancer types in humans and its incidence is on the rise, especially in countries where the ozone layer is thinning. The correct and timely diagnosis of suspicious skin lesions is one of the most important factors in the therapeutical outcome.

At present most dermatologists rely on their experience of visual assessment to distinguish benign and malign skin lesions [1] like pigmented nevi, seborrhoeic keratosis or basal cell carcinoma and malignant melanoma, as well as requiring biopsies of the affected skin.

To complicate matters, Cutaneous T-Cell Lymphoma (CTCL) is a blood cancer type with symptoms that are exhibited as skin lesions as well. Again a timely diagnosis and staging is very crucial for a successful treatment [2].

New technologies to assist in identifying and diagnosing skin lesion and to minimize invasive biopsies have been developed, like hand-held magnification devices and computer-aided image analysis. Colourd image processing methods were introduced for melanoma [3] which focused on non-constant visual information of skin lesions. Neural network diagnosis of skin lesion has previously been applied by classifying extracted features from digitized dermoscopy images of lesions [4][5]. The extracted features are based on geometry, colors, and texture of the lesions, involving complex image processing techniques. Recently with Raman spectroscopy

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 411–418, 2010.

the molecular structure of skin lesions are exploited [6], but due to harmful effect of laser beam on sensitive skin surface, is least in practice for dermatologist. Optical spectroscopy is another technology that is being established to aid in skin lesion diagnosis [7], as the multi-spectral nature of this imaging method allows to detect and classify multiple physiological changes like those associated with increased vasculature, cellular structure, oxygen consumption or edema in tumors [8].

We propose a computer-aided system using optical spectroscopy that keeps track of the progression of skin lesions and assists in quantification and classification of skin diseases in order to assist dermatologists in the decision making process. In this paper we present a framework for acquiring spectroscopic data of skin lesions and classifying them using support vector machines (SVM). We report on the classification results obtained from optical spectroscopy using various skin lesions of 48 patients.

In the following section we describe the materials and methods used in this study and the last section explains experiments and their results.

2 Materials and Methods

Here we describe the employed algorithms and the instrument used for data acquisition.

2.1 System Setup

The system contains a hand–held reflectance spectroscopy probe (StellarNet Inc., Oldsmar, FL, USA), consisting of 6×200µm illumination fibers arrayed around one 600µm acquisition fiber. The probe has an infrared optical tracking target attached in order to be able to determine its position and orientation in real–time, see fig. 1(a). The tracking system consists of four ARTtrack2 infrared cameras (A.R.T. GmbH, Weilheim, Germany) positioned to track within a volume of 2×2×2 [m^3].

A 178−1132 [nm], 2048 [pix], 12bit CCD spectrometer (StellarNet Inc., Oldsmar, FL, USA) is connected to the acquisition fiber, and a 12 [W] tungsten lamp connected to the illumination fibers as a light source. The spectrometer is controlled by a data processing unit to acquire spectra synchronously with the tracking information of the probe. The data-processing unit is used to run the application that handles the incoming data (spectral and spatial) and the classification. An overview of the entire setup is displayed in fig. 1 (b).

2.2 Data Processing

The spectral data is acquired as a 2048 vector of the floating points values $x_i \in R^{2048}, i = 1,\ldots,n$ where n denotes the number of measurements. Each x_i represents the discretized reflective spectrum from 178 [nm] to 1132 [nm] (due to limitation of hardware)of the ith measurement and is stored normalized as

$$\hat{x}_i = \frac{x_i}{\|x_i\|_2} \qquad \text{where } i = 1\ldots n \tag{1}$$

To reduce the dimensions of the input data, principal components analysis (PCA) is applied. The resulting spectrum of eigenvalues $(e_j^i)_{j=1,\ldots,2048}$ is sorted descending by magnitude. Since the highest eigenvalues represent the most relevant components, a cut-off value C_{PCA} is chosen, such that the final input data y_i for the classification algorithm from measurement x_i ($i = 1,\ldots,n$) is

$$y_i = (e_j^i)_{j=1,\ldots,C_{PCA}}. \tag{2}$$

The cut-off value C_{PCA} is chosen empirically from the data. Fig. 2 is showing a representative example of $(e_j^i)_{j=1,\ldots,2048}$ from which C_{PCA} was selected as one of {2,3,4,5}.

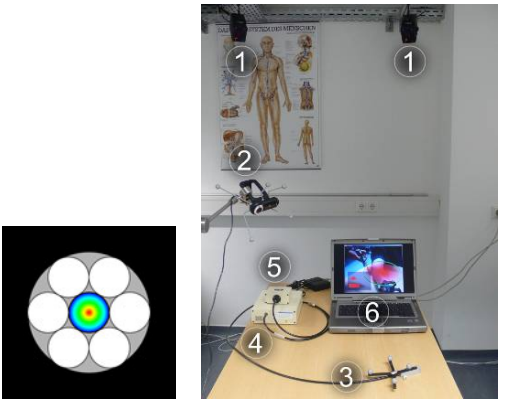


Fig. 1. (a) Schematic of the fiber arrangement in the spectroscopy probe: 6×200µm illumination fibers arrayed around one 600µm acquisition fiber. **(b)** System setup: (1) tracking cameras, (2) regular camera (for augmented reality visualization), (3) tracked probe, (4) spectrometer, (5) light source, and (6) data–processing unit.

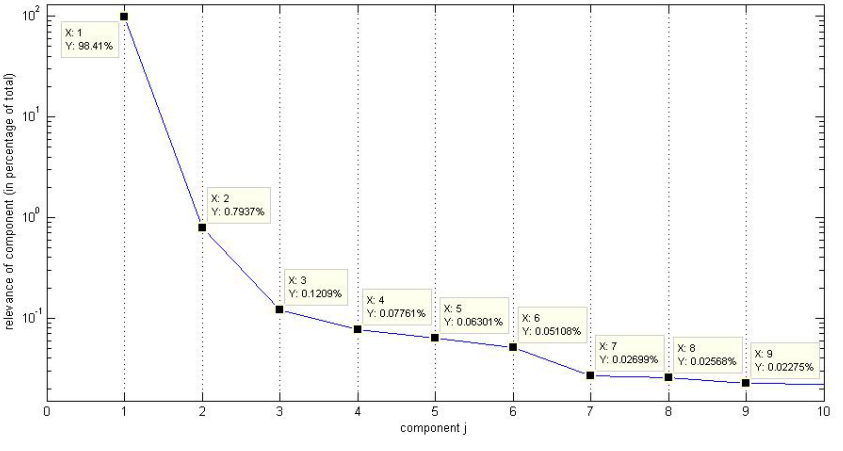


Fig. 2. Representative example of the first part of the sorted PCA eigenvalue spectrum (e_j^i), the y-axis shows the values of the component as a percentage of the total in log scale

2.3 Classification

Classification is performed by a support vector machine (SVM), [9]. SVM was selected as the method of choice as it allows to linearly classify data in a high-dimensional feature space that is non-linearly related to the input space via the use of specific kernel functions, such as polynomial functions or radial basis functions (RBF). This way we can build complex enough models for skin lesion classification while still being able to compute directly in the input space.

The SVM classifier needs to be trained first before using it, thus we partition our (already reduced) input data (y_i), $i=1,\ldots,n$ into two partitions, $T \subset \{1,\ldots,n\}$ the training set and $V \subset \{1,\ldots,n\}$ the testing (or validation) set with $T \cup V = \{1,\ldots,n\}$ and $T \cap V = \{\}$. The training data set T is labeled manually into two classes with the ground truth, $l(y_i)=\pm 1$. Once the classifier is trained, a simple evaluation of the decision function $d(y_i)= \pm 1$ will yield the classification of any data y_i.

In detail, SVM is trying to separate the data $\phi(y_i)$ mapped by the selected kernel function ϕ by a hyperplane $w^T \phi(y_i)+b=0$ with w the normal vector and b the translation. The decision function then is $d(y_i) = \mathrm{sgn}(w^T \phi(y_i)+b)$. Maximizing the margin and introducing slack variables $\xi = (\xi_i)$ for non-separable data, we receive the primal optimization problem:

$$\begin{aligned} & \min_{w,b,\xi} = \frac{1}{2} w^T w + C \sum_{i \in T} \xi_i \\ \textit{with constrains} \quad & l(y_i)(w^T \phi(y_i) + b) \geq 1 - \xi_i \\ & \xi_i \geq 0 \quad \textit{for } i \in T, \end{aligned} \tag{3}$$

where C is a user-determined penalty parameter. Switching to the dual optimization problem allows for easier computation,

$$\begin{aligned} & \min_{\alpha} = \frac{1}{2} \alpha^T Q \alpha - e^T \alpha \\ \textit{with constrains} \quad & 0 \leq \alpha_i \leq C \quad \textit{for } i \in T \\ & \sum_{i \in T} y_i \alpha_i = 0, \end{aligned} \tag{4}$$

where $\alpha = (\alpha_i)$ are the so-called support vectors, $e = [1,\ldots,1]^T$ and Q is the positive semidefinite matrix formed by $Q_{jk} = l(y_j) l(y_k) K(y_j, y_k)$, and $K(y_j, y_k) = \phi(y_j)^T \phi(y_k)$ is the kernel function built from ϕ. Once this optimization problem is solved, we determine the hyperplane parameters w and b, w directly as $w = \sum_{i \in T} \alpha_i \, l(y_i) \, \phi(y_i)$ and b via one of the Karush-Kuhn-Tucker conditions as $b = -l(y_i) y_i^T w$, for those i with $0 < \alpha_i < C$. Thus the decision function of the trained SVM classifier ends up as

$$d(y_i) = sgn(w^T \phi(y_i) + b) = sgn\left(\sum_{j \in T} \alpha_j \, l(y_j) K(y_j, y_i) + b \right). \tag{5}$$

3 Experiments and Results

3.1 Experiments

We collected 436 spectroscopic data points (x_i), i=1,…,436 from the skin of 48 patients, 326 measurements were of skin lesions, 110 measurements were of normal skin. Exemplary picture of the lesion imaged is shown in fig. 4.

All data was manually labeled into the two classes normal skin $l(x_i)$=1 and lesion $l(x_i)$=-1. The 436 data points were randomly separated into a training data set T and a testing (validation) data set V with $|T|$=305 and $|V|$=131, however retaining the

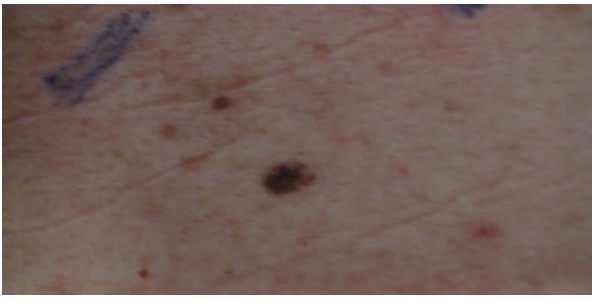


Fig. 4. Skin moles

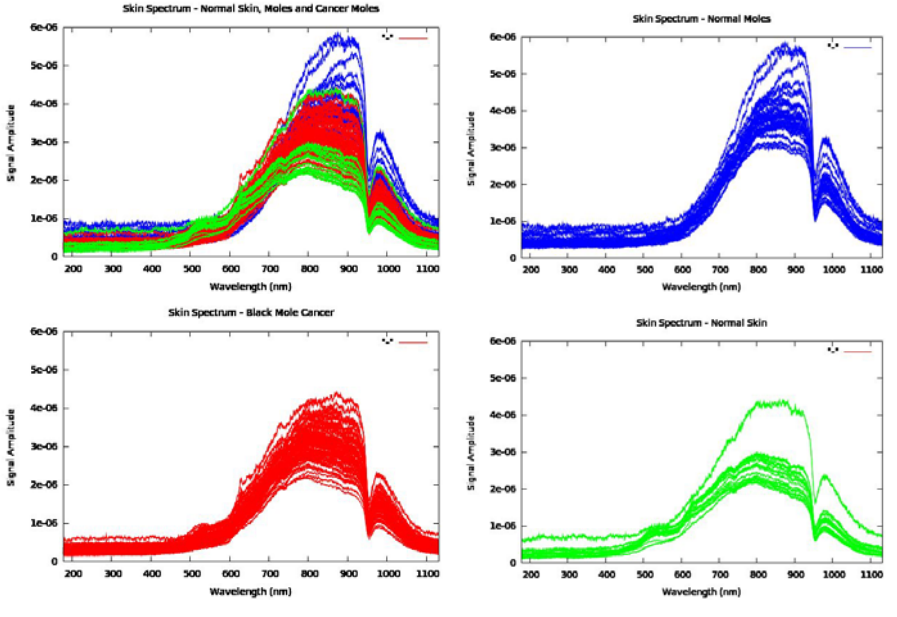


Fig. 5. Graph of all normalized spectra $\hat{x}_i$ from the training data set T, color-coded as blue for skin moles, red mole cancer and green for normal skin.

balance of both sets containing 50% each of the two classes. A color-coded representation of the normalized skin spectra $\hat{x}_i$, $i \in T$, of the training data set T is shown in fig. 5.

Before classification, PCA was applied to the $\hat{x}_i$ for dimension reduction to yield our classification input y_i. The eigenvalue cut-off C_{PCA} was empirically chosen as one of $C_{PCA} \in \{2,3,4,5\}$.

The SVM classifier (we used C-SVM from LibSVM, [10]) was then trained using the training data set T. As there are multiple parameters to be selected, like for example the penalty parameter C, we performed a cross-validation of 3 folds via parallel grid search. The average accuracy on the prediction of the validation fold is the cross validation accuracy.

3.2. Results

The cross-validation of the training data set T determined, among others, the parameters C=-5 and γ=-7. For the further parameters C_{PCA} and the choice of the kernel (linear, polynomial, radial basis function (RBF) or sigmoid) we performed cross validation of the training data set T, the results are shown in Table 1. The best results were received consistently by using the RBF kernel, while for C_{PCA} the value of 5 turned out to be the best choice with an accuracy of 97±8.3.

Table 1. Results of the cross-validation using the training dataset T

C-SVC Training				
	Linear Kernel	Poly Kernel	RBF Kernel	Sigmoid Kernel
$C_{PCA} = 2$	95±9.2	96±8.3	95±7.5	95±10.1
$C_{PCA} = 3$	95±8.3	96±6.7	97±9.5	96±10.5
$C_{PCA} = 4$	95±11.5	97±7.2	97±8.7	96±8.6
$C_{PCA} = 5$	96±9.2	97±10.5	97±8.3	97±7.7

With the training of the classifier completed, we studied the accuracy of the testing (validation) data set V. We compared the manual ground truth labeling $l(yi)$ for data point y_i with the computed decision function $d(y_i)$ to compute the accuracy as follows

$$Accuracy = \frac{\#\,of\ \ correctly\ \ \ predicted\ \ \ data}{\#\,total\ \ data} \times 100\,\% = \frac{\left|\{i | l(y_i) d(y_i) > 0\}\right|}{|V|} \times 100\,\% \qquad (6)$$

The results are shown in Table 2. We achieve the same accuracy of 94.9% for the kernels RBF the C_{PCA} values of 4 and 5. This corresponds to fig. 2, where it is clear that between C_{PCA} 4 and 5 there is only very little difference. In total we received the best results using the RBF kernel and C_{PCA}=5.

Table 2. Classification accuracy results using the testing dataset *V*

Testing				
	Linear Kernel	Poly Kernel	RBF Kernel	Sigmoid Kernel
$C_{PCA} = 2$	86.8%	90.3%	89.9%	88.8%
$C_{PCA} = 3$	89.3%	92.5%	91.8%	90.3%
$C_{PCA} = 4$	91.9%	92.9%	94.9%	94.1%
$C_{PCA} = 5$	92.1%	93.6%	94.9%	94.6%

4 Conclusion

We present a portable, affordable setup for optical spectroscopy and SVM-based classification of skin lesions. Our experiments on patient's data served as a base to choose and tune the various parameters in the classification chain. The results of 94.9% accuracy in distinguishing normal skin from any type of skin lesion are comparable to those of a well-trained dermatologist using visual inspection [11].

Future work includes considering some factors which have not been addressed yet, such as the influence of external light and obstruction by hair (all measurements avoided hair) and comparison with other techniques such as neural networks and manifold learning. The next step then is to established complete framework for clinical evaluation in dermatology department and its determined parameters to classify different skin lesions of diseased patients. These results promise that computer-assisted multi-spectral imaging and classification is the path into the future for dermatological screening.

References

1. Lindelöf, B., Hedblad, M.: Accuracy in the clinical diagnosis and pattern of malignant melanoma at a dermatological clinic. J. Dermatol. 21(7), 461–464 (1994)
2. Klemke, C.D., Goerdt, S., Schrama, D., Becker, J.C.: New insights into the molecular biology and targeted therapy of cutaneous t-cell lymphomas. Journal der Deutschen Dermatologischen Gesellschaft 4(5), 395–406 (2006)
3. Colot, O., Devinoy, R., Sombo, A., de Brucq, D.: A color image processing method for melanoma detection. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, p. 562. Springer, Heidelberg (1998)
4. Hintz-Madsen, M., Hansen, L., Larsen, J., Drzewiecki, K.: A probabilistic neural network framework for detection of malignant melanoma. In: Naguib, R., Sherbet, G. (eds.) Artificial Neural Networks in Cancer Diagnosis, Prognosis and Patient Management, pp. 141–183. CRC, Boca Raton (2001)
5. Rubegni, P., Burroni, M., Cevenini, G., Perotti, R., Dell'Eva, G., Barbini, P., Fimiani, M., Andreassi, L.: Digital dermoscopy analysis and artificial neural network for the differentiation of clinically atypical pigmented skin lesions: A retrospective study. J. Investigat. Dermatol. 119, 471–474 (2002)

6. Sigurdsson, S., Philipsen, P., Hansen, L., Larsen, J., Gniadecka, M., Wulf, H.: Detection of Skin Cancer by Classification of Raman Spectra. IEEE Transactions on Biomedical Engineering 51(10) (2004)
7. McIntosh, L.M., Summers, R., Jackson, M., Mantsch, H.H., Mansfield, J.R., Howlett, M., Crowson, A.N., Toole, J.W.: Towards non-invasive screening of skin lesions by near-infrared spectroscopy. J. Invest. Dermatol. 116(1), 175–181 (2001)
8. Kim, A., Kasthuri, U., Wilson, B.C., White, A., Martel, A.L.: Preliminary Clinical Results for the In Vivo Detection of Breast Cancer Using Interstitial Diffuse Optical Spectroscopy. In: MICCAI Workshop on Biophotonics Imaging for Diagnostics and Treatment, p. 75 (2006), ISSN 1601-2321
9. Cristianini, N., Shawe-Taylor, J.: An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press, Cambridge (2000), ISBN:0521780195
10. Chang, C.-C., Lin, C.-J.: LIBSVM: A library for support vector machines (2001), `http://www.csie.ntu.edu.tw/~cjlin/libsvm`
11. Morton, C.A., Mackie, R.M.: Clinical accuracy of the diagnosis of cutaneous malignant melanoma Br. J. Dermatol. 138, 283–287 (1998)

Segmentation of Vertebral Bodies in MR Images Based on Geometrical Models in 3D

Darko Štern, Boštjan Likar, Franjo Pernuš, and Tomaž Vrtovec

Faculty of Electrical Engineering, University of Ljubljana,
Tržaška cesta 25, SI-1000 Ljubljana, Slovenia
{darko.stern,bostjan.likar,franjo.pernus,
tomaz.vrtovec}@fe. uni-lj.si

Abstract. Segmentation of vertebrae provides means for reliable measurement of vertebral deformations, which is important for the diagnosis and therapy of pathological conditions affecting the spine. In this paper we propose a method for segmentation of vertebral bodies in three-dimensional (3D) magnetic resonance (MR) images that is based on efficient geometrical modeling of the vertebral body and evaluation of dissimilarity between the vertebral body and the surrounding soft tissue in MR images. The results show that the proposed geometrical model of the vertebral body can describe a variety of vertebral body shapes and therefore the method may be used for quantitative assessment of vertebral body deformations or initialization of whole vertebra segmentation.

Keywords: magnetic resonance (MR), spine, vertebral body, shape, modeling, segmentation.

1 Introduction

Pathological conditions affecting the spine, such as low back pain, instability, osteoporosis, stenosis, spondylolisthesis or herniations, have become acute problems of modern society. Among the techniques for analysis of spine images, segmentation stands out as a technique that may not only improve visualization of vertebrae, but also provide measurements of vertebral deformations that may aid clinical diagnostics and therapy.

Several methods for segmentation of vertebrae and vertebral structures have been proposed. Starting from a point in the center of each vertebral body, Mastmeyer et al. [1] segmented vertebrae in computer tomography (CT) images by a combination of viscous deformable models and evaluated the geometrical shape of the vertebral body. Kim et al. [2] automatically segmented vertebrae using a region growing method inside a volume limited by a three-dimensional (3D) fence that was obtained by a deformable model. By applying statistical models of shape, gradient and appearance of the spinal structure in 3D, Klinder et al. [3] detected, identified and segmented the vertebrae in CT images. To segment vertebrae in magnetic resonance (MR) images, Peng et al. [4] used template-based matching and edge connectivity tracing procedure in a manually selected sagittal cross-section. Normalized-cut segmentation algorithm,

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 419–428, 2010.

proposed by Huang et al. [5], was also performed in a single sagittal cross-section. Methods for segmentation of vertebrae in a single cross-section usually presume pre-selection of a plane in which all structures of interest are visible. Pre-selection of such plane may be too demanding or even not feasible, as in the case of large vertebral deformations. Above all, the main disadvantage of vertebra segmentation in a single cross-section is its inability to capture the complexity of the 3D vertebral anatomy. Hoad et al. [6] proposed a method for segmentation of vertebral bodies in 3D MR images. However, the segmentation was based on fitting an ellipse in a single axial cross-section and then connecting the obtained ellipses into a volume. To the best of our knowledge, there is no method for segmentation of vertebrae in MR images that is performed completely in 3D.

In this paper we propose a method for segmentation of vertebral bodies in 3D MR images that is based on efficient modeling of the vertebral body shape. The segmentation is performed by optimizing the parameters of a geometrical model so that it is aligned with the vertebral body in the image. To obtain the parameters of the geometrical model, we exploit the dissimilarity between the appearance of the bone structure and the soft tissues that surround the vertebral body in MR images. The preliminary results show that the optimization criterion can be used for segmentation of vertebral bodies and the geometrical model can describe various vertebral body shapes. The proposed method may be therefore used for initializing whole vertebra segmentation or quantitative description of vertebral body deformations.

2 Method

2.1 Geometrical Model of the Vertebral Body

Our initial hypothesis is that the vertebral body can be modeled with an elliptical cylinder in 3D. The following nine parameters describe such geometrical model: x_0, y_0 and z_0 define the location of the center, a, b and h define the semi-major axis, the semi-minor axis and the height, and α, β and γ define the inclination of the elliptical cylinder in 3D. The elliptical cylinder $F(x, y, z)$ can be presented as an implicit function of a superellipsoid [7]:

$$F(x,y,z)=\left(\frac{x^2+y^2}{r(\theta)^2}\right)^{\frac{1}{\varepsilon_1}}+\left(\frac{z}{h}\right)^{\frac{1}{\varepsilon_1}}, \tag{1}$$

where the parameter ε_1 defines the curvature of the superellipsoid edges and, to obtain sharp edges, is set to 0.1 (Fig. 1a). The function of the ellipse radius $r(\theta)$ in the polar coordinate system:

$$r(\theta)=\frac{a\cdot b}{\sqrt{(a\cdot\cos\theta)^2+(b\cdot\sin\theta)^2}}, \tag{2}$$

defines the shape of the basis of the superellipsoid, where $\theta = \arctan(y/x)$ is the radius angle measured from the major axis.

To change the elliptical cylinder into a more accurate description of the vertebral body, the elliptic shape was modified by superimposing four Gaussian functions, centered at angles θ_i and with standard deviations σ_i; $i = 1, \ldots, 4$, to the ellipse radius. The resulting changes in the shape of the elliptical cylinder allow modeling of the vertebral body at the location of vertebral pedicles and the vertebral foramen (Fig. 1b). The function of the ellipse radius $r(\theta)$ in Eq. 1 is therefore modified into $R(\theta)$:

$$R(\theta) = r(\theta) \cdot \left[1 + m_1 \cdot e^{-\frac{\theta-\theta_1}{\sigma_1}} + m_2 \cdot e^{-\frac{\theta-\theta_2}{\sigma_2}} + m_3 \cdot e^{-\frac{\theta-\theta_3}{\sigma_3}} + m_4 \cdot e^{-\frac{\theta-\theta_4}{\sigma_4}} \right], \quad (3)$$

where the weight m_i; $i = 1,..,4$ defines the magnitude of the modification caused by i^{th} Gaussian function.

The wall of normal and especially pathological vertebrae is usually curved inward. To model the curvature of the vertebral body wall, we use a cosine function with the period equal to the height h of the cylinder and amplitude c_W. The wall of the vertebral foramen may also be curved inward, which we model with another cosine function with amplitude c_F. The amplitudes c_W and c_F are considered as parameters of the geometrical model (Fig. 1c).

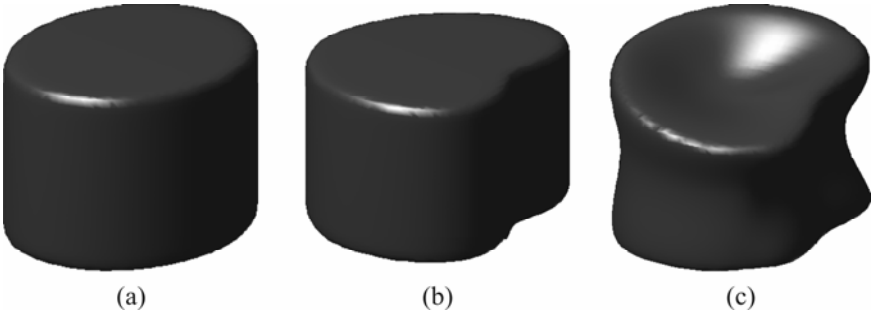


Fig. 1. Geometrical model of the vertebral body generated with (a) an elliptical cylinder, (b) a cylinder with modified basis and (c) a cylinder modified with all proposed parameters

The endplates of normal vertebrae are usually parallel but a variety of conditions (such as osteoporosis, severe trauma, congenital deformities, Scheuermann's disease, osteoarthritis or multiple myeloma) may deform their shape. Moreover, endplate shape is usually concave. We model such shape variations separately for the superior and inferior endplate with two-dimensional (2D) cosine functions with the period equal to the radius of the cylinder basis (Eq. 3) and amplitude c_S and c_I (for superior and inferior endplate, respectively) (Fig 1c). The wedge and crush deformity are deformations where the endplates are inclined toward the anterior or posterior part of the vertebral body [8]. As the inclinations of the superior and inferior endplates are independent, we model them separately with magnitudes s_S and s_I and angles ψ_S and ψ_I (for superior and inferior endplate, respectively) (Fig. 1c).

To summarize, the following 29 parameters are used to generate the geometrical model of the vertebral body: x_0, y_0 and z_0 (center of the vertebral body); α, β and γ

(inclination of the vertebral body); a, b and h (parameters of the elliptical cylinder); θ_i, σ_i and m_i (parameters of the superimposed Gaussian functions to the ellipse in the cylinder basis; $i = 1, ..., 4$); c_W and c_F (curvature of the vertebral body wall and foramen); c_S and c_I (concavity of the superior and inferior endplate); s_S, s_I, ψ_S and ψ_I (magnitude and angle of the inclination of the superior and inferior endplate).

2.2 Segmentation of the Vertebral Body

The segmentation of the vertebral body in 3D MR images is performed by optimizing the parameters of the geometrical model to achieve the best match of the model to the vertebral body in the image. By observing the appearance of the vertebral body and surrounding soft tissues in MR images, the following conclusions can be made:

- image intensities that belong to the vertebral body are different from the intensities that belong to the surrounding soft tissues, and
- the vertebral body as a homogeneous structure has a small standard deviation of image intensities, while the image intensity gradients are concentrated around the edges of the vertebral body.

Basing on these conclusions, the matching of the geometrical model to the vertebral body is performed by maximizing the dissimilarity between image features within the model and features within the area that surrounds the model. The dissimilarity measure can be represented by:

$$DM = \frac{\sum_j g_{surr}(j)}{\sigma^I_{\text{mod}}} \cdot \left| p^I_{\text{mod}} - p^I_{surr} \right|, \tag{4}$$

where $g_{surr}(j)$ is the gradient magnitude at point j that is within the area that surrounds the model, σ^I_{mod} is the standard deviation of image intensities within the model, and p^I_{mod} and p^I_{surr} are the probability distributions of image intensities within the model and within the area that surrounds the model, respectively. The width K of the area that surrounds the model is initially large in order to capture the global position of the vertebral body in the image and is then gradually reduced in order to capture only the soft tissues around the edge of the vertebral body.

3 Experiments

3.1 Images, Ground Truth and Method Evaluation

The performance of the proposed method was evaluated on six T_2-weighted MR images of lumbar spine containing 30 vertebrae from L1 to L5. Three were axial MR images with voxel size $0.4 \times 0.4 \times 3$ mm, while three were sagittal MR images with voxel size $1.0 \times 1.0 \times 1.0$ mm. For the purpose of quantitative evaluation of the proposed method, the ground truth was determined for each vertebra by manually identified anatomical points. A single observer determined 16 characteristical points in the mid-axial, mid-sagittal and mid-coronal plane of each vertebra (Fig. 2). Based

on the identified points, the center and inclination of the vertebral body were calculated.

The performance of the method was measured quantitatively by computing the radial Euclidian distance between the ground truth points and the obtained model of the vertebral body [7]. The radial Euclidian distance is defined as the distance between an arbitrary point and the model along the line that passes through the point and the center of the model. Based on the implicit function $F(x, y, z)$ (Eq. 1), the radial Euclidean distance of the model from the ground truth point (x_G, y_G, z_G) is:

$$d = d_G \cdot \left| 1 - F^{-\frac{0.1}{2}}\left(x_G, y_G, z_G\right) \right|, \tag{5}$$

where d_G is the Euclidian distance of the point from the center of the model.

To evaluate the performance of the proposed method against different initial positions of the geometrical model, we initialized the center (x_0, y_0 and z_0) and inclination (α, β and γ) of the geometrical model in randomly generated displacements from their ground truth values. The displacements were uniformly distributed and ranged up to 10 mm, where the inclination angles were transformed into translations so that volume displacement was preserved. For each vertebra, 100 initial displacements were applied, resulting in a total of 3000 displacements for all vertebrae. The initial shape of each vertebral body was represented by an elliptical cylinder with the major axis, minor axis and height set to the width, length and height, respectively, of the average lumbar vertebral body [9].

3.2 Implementation Details

For the purpose of vertebral body segmentation in 3D, all images were resampled to anisotropic voxel resolution of $1.0 \times 1.0 \times 1.0$ mm using trilinear interpolation. To reduce image noise and local inhomogeneity, the images were blurred with a 3D Gaussian filter (standard deviation $\sigma = 1$ mm). The gradient magnitudes were calculated using the 3D Sobel gradient operator (kernel size $3 \times 3 \times 3$ mm) [10].

Starting from a single, user defined point located near the center of the vertebral body, the segmentation was performed by optimizing the parameters of the geometrical model of the vertebral body with the simplex algorithm [11]. The initial shape of each vertebral body was represented by an elliptical cylinder generated basing on the average width ($2a = 45$ mm), length ($2b = 30$ mm) and height ($2h = 24$ mm) of the lumbar vertebral body [9]. The optimization was carried out in three steps. In the first step, nine parameters of the vertebral body model (x_0, y_0, z_0, α, β, γ, a, b and h) were obtained by translating, rotating and scaling the original elliptical cylinder, with the thickness of the surrounding area of the model set to $K = 8$ mm. The obtained parameters were used to generate the starting model in the second step, where 12 new parameters (θ_i, σ_i and m_i; $i = 1,.., 4$) were added in optimization and the surrounding area was reduced to $K = 6$ mm. The initial angles of Gaussian functions close to the pedicles, foramen and interior part of the vertebral body were equal to $\theta_1 = 45°$, $\theta_2 = -45°$, $\theta_3 = 0°$ and $\theta_4 = 180°$, respectively, from the minor axis of the obtained ellipse, while the parameters σ_i and m_i; $i = 1, ..., 4$; were initialized to $\pi/8$ and 0.3, respectively. In the third step of the optimization, the remaining eight parameters

(c_W, c_F, c_T, c_B, s_T, s_B, ψ_T and ψ_B) were added to obtain the local characteristics of the vertebral body (e.g. curvature of the vertebral body wall, curvature of the foramen, concavity and inclination of the endplates) and the surrounding area of the model was reduced to $K = 4$ mm.

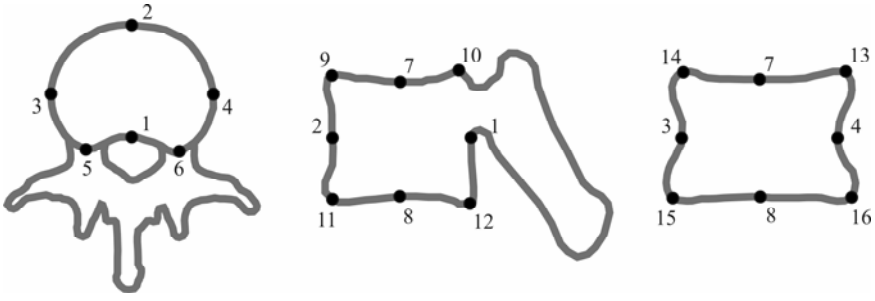


Fig. 2. The ground truth points were determined in the mid-axial, mid-sagittal and mid-coronal cross-section of each vertebra

4 Results

The proposed method for segmentation of vertebral bodies in 3D was successfully applied to T_2-weighted MR images of lumbar spine. The results of the experiments are presented as radial Euclidian distances (Eq. 5) from the ground truth points. For each vertebra, we present the mean difference and standard deviation between the model and the ground truth points for successful segmentations after each optimization step (Table 1). A segmentation was considered successful if the mean distance of the obtained model from the ground truth points was less than 5 mm. The overall mean difference (± standard deviation) for successful segmentations of all vertebrae after the first optimization step (i.e. when nine parameters of the vertebral body were optimized) was 2.8 ± 1.9 mm (success rate 77.5%), after the second optimization step (i.e. when optimizing 21 parameters) 2.7 ± 1.6 (success rate 81.0%), and after the third optimization step (i.e. when optimizing all 29 parameters) 2.1 ± 1.3 mm (success rate 86.7%). Figure 3 shows the scatter diagrams of the displacements of the initial model and the distance between the obtained model and the ground truth points. Detailed results for selected vertebra are presented in figure 4.

5 Discussion and Conclusion

A method for segmentation of vertebral bodies in 3D MR images based on a geometrical model in 3D was proposed. The main advantage of geometrical modeling over the widespread statistical modeling is that it is not constrained to the deformations included in the training data set. Statistical modeling also requires initialization near the final solution, while in the proposed method, initialization was represented by a single point near the center of the vertebral body. Moreover, the capture range of the proposed method is quite large, since for the initial displacement

of up to 10 mm, 86.7% of vertebral bodies were successfully segmented. As the method is initialized with a single point and automatic methods for the detection of the centers of vertebral bodies already exist [12], we can assume that the proposed method for segmentation of vertebral bodies in MR images performs automatically.

Table 1. Mean difference Δ (± standard deviation σ) between the model and the ground truth points for successful segmentations after each optimization step, reported for all lumbar vertebrae used in the experiment. The success rate is defined by the number of segmentations for which the mean distance between the model and the ground truth points is less than 5 mm.

Vertebrae		Spine 1	Spine 2	Spine 3	Spine 4	Spine 5	Spine 6
L1	$\Delta_1 \pm \sigma$ (mm)	3.4±2.4	2.0±1.4	3.6±2.9	3.8±3.4	3.9±3.0	3.3±1.7
	$\Delta_2 \pm \sigma$ (mm)	3.1±2.1	2.3±1.3	3.3±2.8	3.7±3.7	3.5±3.5	2.8±1.7
	$\Delta_3 \pm \sigma$ (mm)	2.1±1.3	1.9±1.1	2.1±1.3	2.0±1.4	2.1±1.4	2.4±1.4
	success rate	92.5%	99.5%	80.6%	78.5%	81.8%	93.5%
L2	$\Delta_1 \pm \sigma$ (mm)	2.9±2.4	2.5±1.7	3.6±2.6	3.1±2.5	3.4±3.0	3.2±1.9
	$\Delta_2 \pm \sigma$ (mm)	2.5±1.8	2.7±1.4	3.2±2.5	3.3±2.4	3.4±3.0	3.0±1.5
	$\Delta_3 \pm \sigma$ (mm)	2.0±1.4	2.3±1.2	2.3±1.3	2.1±1.5	2.4±1.1	2.4±1.4
	success rate	97.5%	98.6%	87.5%	83.4%	87.1%	91.1%
L3	$\Delta_1 \pm \sigma$ (mm)	3.6±2.8	2.9±2.3	3.7±2.7	3.6±2.5	3.7±2.6	2.8±2.0
	$\Delta_2 \pm \sigma$ (mm)	3.1±2.3	2.9±2.0	3.1±2.5	3.7±2.4	3.6±2.4	2.8±1.7
	$\Delta_3 \pm \sigma$ (mm)	1.9±1.4	1.8±1.4	1.9±1.3	2.1±1.4	2.1±1.2	2.1±1.3
	success rate	85.4%	88.5%	81.9%	77.2%	77.2%	94.8%
L4	$\Delta_1 \pm \sigma$ (mm)	3.1±2.6	3.0±2.0	4.6±3.4	3.9±2.5	3.0±3.4	2.5±1.7
	$\Delta_2 \pm \sigma$ (mm)	3.4±2.3	2.9±1.6	4.1±2.8	3.9±2.1	3.0±3.3	2.3±1.2
	$\Delta_3 \pm \sigma$ (mm)	2.1±1.5	2.5±1.4	2.0±1.3	2.3±1.4	2.1±1.1	1.6±0.9
	success rate	89.3%	92.3%	78.3%	81.9%	87.2%	91.7%
L5	$\Delta_1 \pm \sigma$ (mm)	2.8±2.9	2.8±1.9	3.7±2.5	3.8±3.0	3.1±2.9	2.1±1.7
	$\Delta_2 \pm \sigma$ (mm)	3.0±2.7	2.6±2.0	3.4±2.4	3.7±2.8	2.9±2.7	2.0±1.1
	$\Delta_3 \pm \sigma$ (mm)	2.3±1.3	2.0±1.4	2.1±1.4	2.3±1.3	2.0±1.4	1.7±1.0
	success rate	83.2%	90.5%	79.3%	77.9%	89.6%	89.1%

To perform successful segmentation based on the geometrical model, the dissimilarity measure that aligns the model to the vertebral body in the image has to be carefully selected (Eq. 4). Namely, the dissimilarity measure that was based only on the intensity difference proved to be highly sensitive to local intensity homogeneity within the vertebral body and caused shrinking of the model toward that homogeneity. On the other hand, when the edges of the vertebral body were weak, the dissimilarity measure based only on gradient information caused the expansion of the model to the surrounding soft tissues. The proposed dissimilarity measure is therefore generated as a measure that weights the shrinking and expanding of the model.

Several studies also assessed the performance of vertebra segmentation quantitatively by measuring the distance from the ground truth. Klinder et al. [3] reported mean distance (± standard deviation) between the obtained surfaces and the

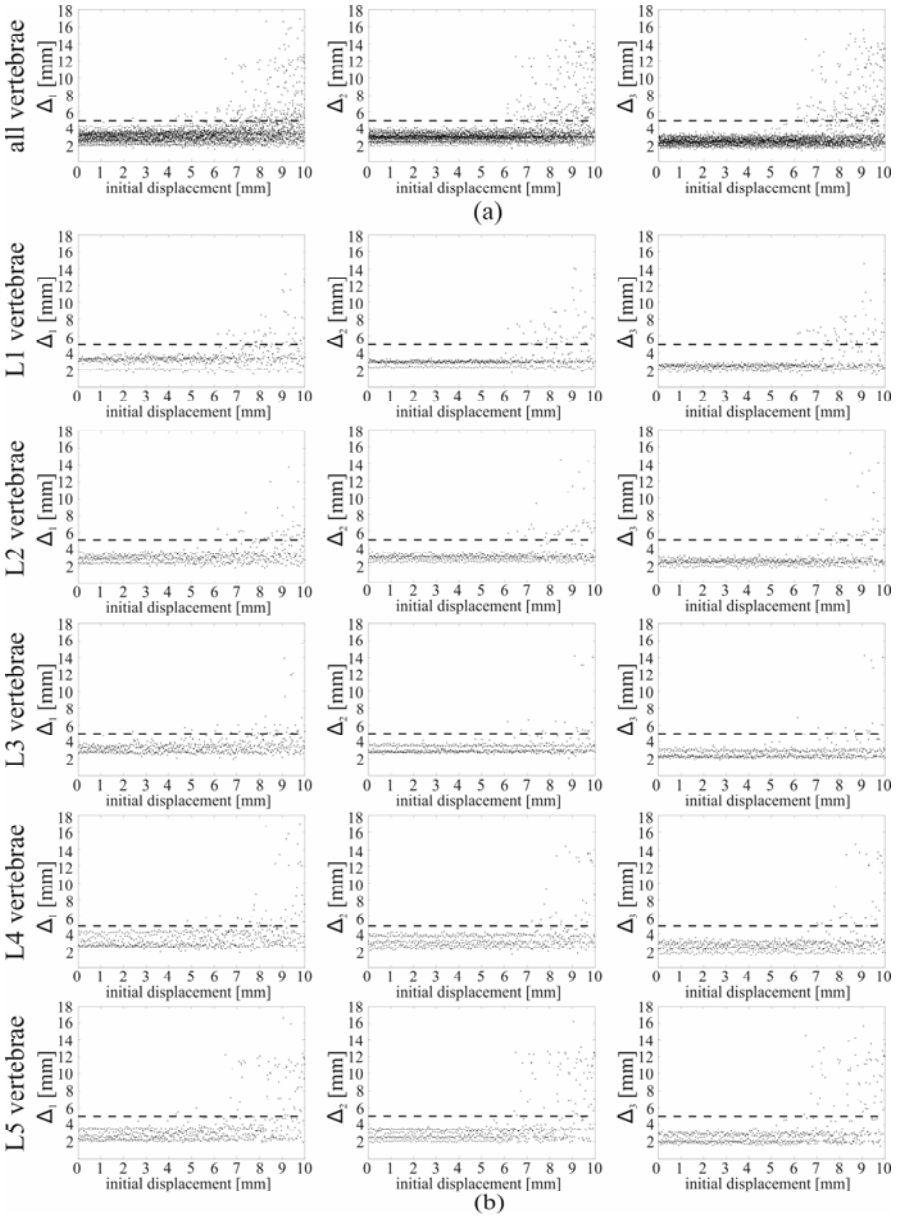


Fig. 3. The scatter diagram of the initial displacement of the vertebral body model versus the distance of the obtained model from the ground truth points after the first optimization step (Δ_1, left), second optimization step (Δ_2 , middle) and third optimization step (Δ_3, right) for (a) all lumbar vertebrae and (b) separately for each lumbar vertebra (L1 to L5).

ground truth mesh grids of 1.12 ± 1.04 mm for whole vertebra segmentation in CT images, while for MR images, Hoad et al. [6] reported 1.12 ± 0.15 mm distance between the obtained and manually segmented vertebrae. By applying the proposed method for segmentation of vertebral bodies in MR images, the distance between the obtained geometrical models and manually identified anatomical points was 2.1 ± 1.3 mm. The preliminary results from 30 vertebrae are therefore comparable to the results reported by other studies. It is also important to note that the ground truth points were determined manually by a single observer. The ground truth should be therefore considered as a valuable guideline that helped us to develop and evaluated

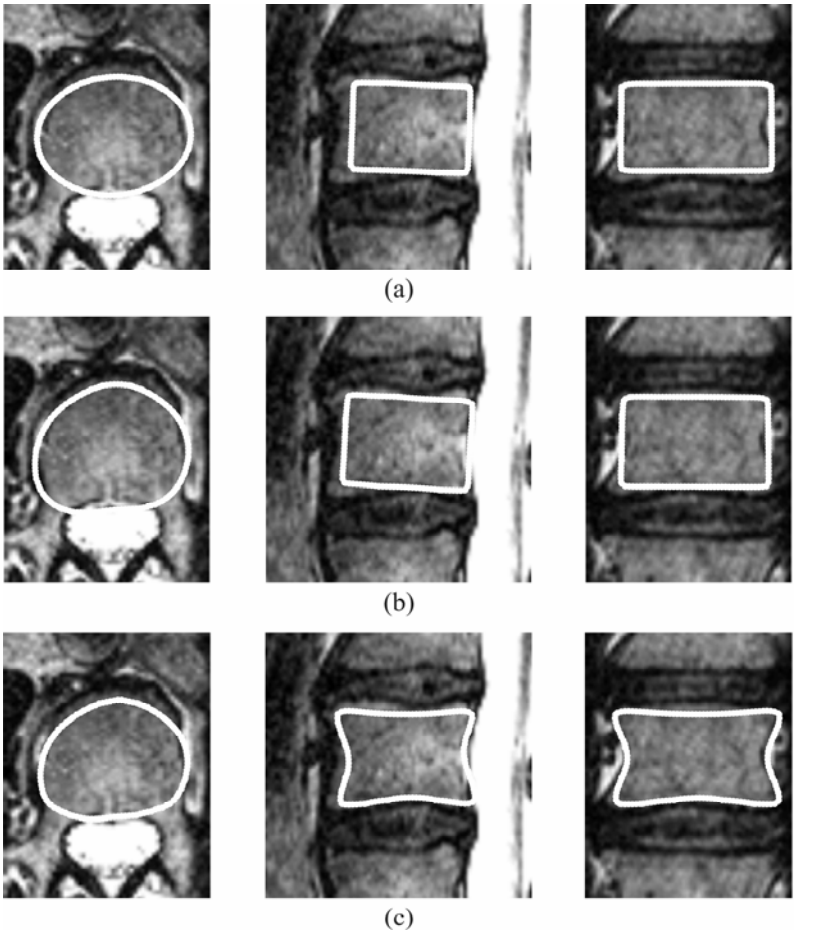


Fig. 4. The segmentation results obtained with (a) 9 parameters, (b) 21 parameters and (c) 29 parameters model of the vertebral body in an axial (left), a sagittal (middle) and a coronal (left) view of the L3 vertebra of the spine 6 (Table 1).

the performance of the proposed method. To obtain a more reliable ground truth, either a manual segmentation of each vertebra has to be performed or the number of manually determined points and observers has to be increased.

Clinical evaluations of vertebral deformities, such as wedge, concave or crush deformity [8], obtained from 2D sagittal or coronal cross-sections do not give a full and accurate interpretation of the deformations in 3D. We propose a method that performs in 3D and the results indicate that 29 parameters of the 3D geometrical model of the vertebral body are sufficient for describing a variety of vertebral body shapes. However, to prove the clinical applicability of the proposed geometrical model, an extensive evaluation of its capability in accurately describing vertebral body deformations has to be performed.

To conclude, we presented a method for segmentation of vertebral bodies in 3D MR images that is based on efficient modeling of the vertebral body shape. The results show that the proposed method can be used for segmentation and that the parameters of the model can describe the geometrical variety of vertebral body shapes. The method may be therefore used for initializing whole vertebra segmentation or quantitative assessment of vertebral body deformations.

References

1. Mastmeyer, A., Engelke, K., Fuchs, C., Kalender, W.A.: A hierarchical 3D segmentation method and the definition of vertebral body coordinate systems for QCT of the lumbar spine. Med. Image Anal. 10, 560–577 (2006)
2. Kim, Y., Kim, D.: A fully automatic vertebra segmentation method using 3D deformable fences. Comput. Med. Imag. Graph. 33, 343–352 (2009)
3. Klinder, T., Ostermann, J., Ehm, M., Franz, A., Kneser, R., Lorenz, C.: Automated model-based vertebra detection, identification, and segmentation in CT images. Med. Image Anal. 13, 471–482 (2009)
4. Peng, Z., Zhong, J., Wee, W., Lee, J.-h.: Automated vertebra detection and segmentation from the whole spine MR images. In: 27th Int. Conf. of the Engineering in Medicine and Biology Society IEEE - EMBC 2005, pp. 2527–2530. IEEE, Piscataway (2005)
5. Huang, S.H., Chu, Y.H., Lai, S.H., Novak, C.L.: Learning-based vertebra detection and iterative normalized-cut segmentation for spinal MRI. IEEE T. Med. Imaging 28, 1595–1605 (2009)
6. Hoad, C.L., Martel, A.L.: Segmentation of MR images for computer-assisted surgery of the lumbar spine. Phys. Med. Biol. 47, 3503–3517 (2002)
7. Jaklič, A., Leonardis, A., Solina, F.: Segmentation and recovery of superquadrics. Kluwer Academic Publishers, Dordrecht (2000)
8. Kurtz, S.M., Edidin, A.A.: Spine technology handbook. Elsevier, Burlington (2006)
9. Masharawi, Y., Salame, K., Mirovsky, Y., Peleg, S., Dar, G., Stemberg, N., Hershkovitz, I.: Vertebral body shape variation in the thoracic and lumbar spine: Characterization of its asymmetry and wedging. Clin. Anat. 21, 46–54 (2008)
10. Gonzalez, R.C., Woods, R.E.: Digital image processing, 3rd edn. Pearson Education, Upper Saddle River (2008)
11. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical recipes in C++, the art of scientific computing, 2nd edn. Cambridge University Press, Cambridge (2002)
12. Štern, D., Likar, B., Pernuš, F., Vrtovec, T.: Automated detection of spinal centerlines, vertebral bodies and intervertebral discs in CT and MR images of lumbar spine. Phys. Med. Biol. 55, 247–264 (2010)

A Learning-Based Approach to Evaluate Registration Success

Christoph Vetter[1,2], Ali Kamen[1],
Parmeshwar Khurd[1], and Rüdiger Westermann[2]

[1] Siemens Corporate Research, Princeton NJ 08544, USA
[2] Computer Science Department, Technische Universität München,
Garching 85748, Germany

Abstract. Clinical trials are more and more relying on medical imaging technologies to quantify changes over time during longitudinal studies. This calls for having an unsupervised batch registration process. However, even good registration algorithms fail, whether that is because of a small capture range, local optima, or because the registration finds an optimum that is not meaningful since the input data contains different anatomical sites. We propose a new method to evaluate the success or failure of batch registrations, so that failed or suspicious registrations can be flagged and manually corrected. The evaluation is based on a support vector machine that evaluates features representing the "goodness" of the registration result. We devise the features to be the distance measured between optima produced by different similarity measures as well as optima resulting from registering subsections of the volumes. The features of 30 volume registrations have been labeled manually and used for the learning phase. Based on a test on unseen 67 volume pairs of varying anatomical sites, we are able to classify 90% of the registrations correctly.

1 Introduction

Longitudinal studies based on modalities like CT (Computed Tomography) reveal disease progression or the efficacy of a certain treatment regimen over time. All these are enabled by an offline registration process. However, even good registration algorithms fail for various reasons. The initial transformation parameters for the registration might lie too far away from the optimal transformation parameters, so that the correct solution is outside the capture range of the algorithm. Another scenario is the presence of local optima in the cost function. Instead of converging to the global optimum, the optimization algorithm might get stuck in such a local optimum. Finally, the registration cannot return meaningful results if the volumes contain different anatomical sites. Unfortunately, it is common in real world applications that data is either incorrectly labeled, or insufficiently labeled, or not labeled at all. For the offline batch registration, it is desirable to use an algorithm that automatically detects suspicious registration results, so that erroneous results can be amended by the user.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 429–437, 2010.

Unfortunately, the intensity-based similarity measures do not readily provide a mechanism to evaluate the result of a registration process. Depending on the domain, the required accuracy varies. Organ movement or tumor growth between consecutive volumes in a time series can be different. Intra-subject registrations have different characteristics than inter-subject registrations. Because of these factors, successful registration results might have a lower similarity value than failed ones. Therefore, we have to rely on independent approaches to assess registration success.

While there has been extensive research on evaluating registration results, most of the work is performed for benchmarking different registration approaches, for the comparison of optimization algorithms, or for the comparison of similarity measures, by comparison to a gold standard [1,2,3]. This does not provide a system to classify the registration outcome of a new pair of datasets. Another branch of work studies the semi-automatic assessment of registration accuracy. One example for this approach is the assessment of the accuracy of registrations between CT images and MR images of the head [4]. The evaluation is performed by segmenting the bone in both MR and CT image and visualizing the overlap. But this method is only applicable to a certain anatomical site and requires some user input. Kybic et al. pursue a more general approach using bootstrap resampling and performing the optimization N times, in order to evaluate the potential error [5,6,7]. This comes at the cost of increased computational complexity by a factor between 10 and 100. The focus is to quantify how accurate the registration result is, detecting complete failure is not addressed. Another example of a statistical approach is presented by Wang et al. [8]. The authors evaluate the accuracy of the registration by statistical confidence intervals based on model fitting. For group based registration, Schestowitz et al. [9,10] provide an accuracy assessment based on the quality of a statistical model of appearance. Ceylan et al. [11] evaluate the consistency of a rigid registration by performing deformable registrations of different subsections of the skull.

In the following, we concentrate on the problem of detecting suspicious registration results for CT data sets from longitudinal studies done for clinical trials. Datasets consist of time points of various CT volumes from different anatomical regions. For the evaluation of the clinical trial, the data sets have to be rigidly registered across the different time points. Then the user is able to perform various steps for the assessment of the outcome of the trial, whether it is segmenting tumors, or performing measurements to assess and quantify changes in organ or tumor. To prevent the registration time from adversely affecting the workflow, the data is preregistered in an offline batch registration step. We want to assess the "goodness" of the pair wise registration. For example, consider a dataset with two time points. Time point 1 contains dataset A (head) and data set B (abdomen+ pelvis), while time point 2 contains data set Y (thorax+abdomen) and dataset Z (pelvis). Registering B to Y and Z results in a meaningful registration, but A cannot be registered to Z (see figure 1). Gross misregistrations like these have to be flagged. A totally accurate registration, however, is neither needed nor possible with rigid registration alone. The higher the accuracy of the

assessment, the less registrations the user has to perform manually, which accelerates the workflow. But even if some misregistrations are not caught, or good registration results are erroneously flagged, the user can correct these mistakes interactively. Since the depicted anatomical sections are not consistent between time points or different trials and we cannot expect accurate labeling of the data content, a novel approach is required to detect gross misregistrations between disparate anatomical sections.

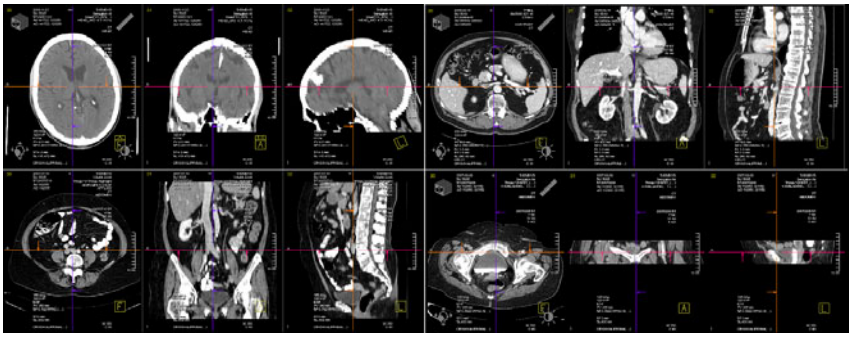


Fig. 1. In this figure, we show examples for typical datasets encountered in our use case. The time point on the left contains head data as well as a scan of pelvis and abdomen, the time point on the right a scan of the thorax and the abdomen as well as a scan of the pelvis.

2 Methods

The evaluation of registrations in this paper is based on the idea of consistency in meaningful registration results. We expect that for a meaningful result, different similarity measures agree closely on the optimum registration parameters, especially in the case of CT registration in which a wide variety of different similarity measures have been employed successfully. However, for registrations that are stuck in some local optimum or have registered disparate parts of the body, the different similarity measures are expected to diverge to different optima that are farther apart. Figure 2 illustrates this concept. Different similarity measures have been used before, in order to achieve a more accurate registration [12], but not specifically for the assessment of success.

A different aspect of the consistency property is spatial consistency of a meaningful registration result (please refer to figure 3). If we consider different regions of the moving volume, all these subdivisions need to have very similar transformation parameters when registered independently, whereas after failed registrations, different subdivisions of the moving volume could end up with different transformation parameters, in the case that they are registered independently of each other. This approach is applicable even if there are not many readily available similarity measures for the problem at hand. Overall, the idea is to

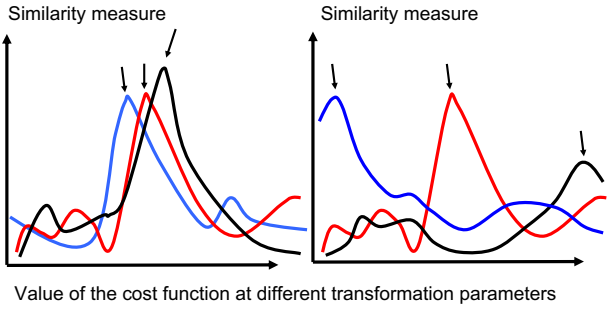


Fig. 2. This figure illustrates the idea of consistency-based evaluation. For a good registration result we expect the optima to be relatively close (left), for a bad registration outcome, we expect the optima further apart (right). The similarity measure is dimensionless, since we are only interested in the optima.

identify consistency properties of a good registration result, and then perform independent registrations and compare the resulting parameters with those of the original registration, expecting a good match for good results and a bad match for bad registration outcomes. A learning algorithm is then used to automatically learn the optimal weights for each feature.

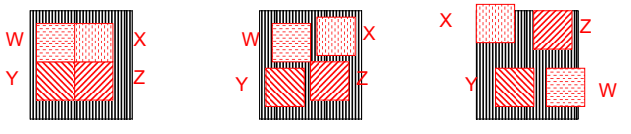


Fig. 3. This figure illustrates the idea of consistency-based evaluation for different subsections. The fixed image is black, the subsections of the moving image are red. The original registration (left) naturally returns identical transformation parameters for all subsections. When registering every subsection independently, we expect little difference to the result of the first registration if the registration is good (middle), whereas the independent registrations might return more divergent results for a bad registration (right). Notice how the spatial relationship between the subsection has not been preserved in the rightmost example.

With these ideas in mind, the workflow of the algorithm is the following. First, a pair of volume data is registered with the registration method of choice. Then a series of independent registrations is performed with a number of different similarity measures and different subsections of the volumes. These independent registrations are initialized with the transformation parameters of the original registration. This allows even similarity measures with smaller capture ranges to converge to the same optimum for a good registration. The differences between the independent registrations to the original registration result are then used as features. In a training step, the features of registrations that have been labeled by a human observer are fed into a learning algorithm. After this training phase,

the same features are then evaluated in order to classify new registration results as failed or successful.

2.1 Features for Registration Quality

For the features based on similarity measures, we use 5 different similarity measures, the sum of squared distances, the sum of absolute distances, cross correlation, correlation ratio, and normalized mutual information [13], as these similarity measures are appropriate for CT-CT registration. For different modalities, a different set of similarity measures has to be chosen. For each of these similarity measures, a coarse registration is performed. The distances between the resulting transformation parameters and the transformation parameters of the original registration for each of the similarity measures are then taken as features. Furthermore, the distance between the two results that are the farthest away from each other is taken as another feature. (In order to not introduce any bias and to remove the likelihood of getting stuck in a local optimum we introduce some jitter to the initialized starting position for the registration and use a different optimization algorithm as well.)

For the subsection-based evaluation, we use the normalized mutual information as similarity measure and divide the moving volume into 8 corners (octants), or divide the moving volume into 4 sections, essentially dividing the volume into stacks of slices. The features that are computed in this case are the minimum distance between the optima of the given registration result and the new registration results for each subsection, the maximum distance, the average distance and the standard deviation of the difference. For consistent registration results, these values can be expected to be small.

Since we are interested in catching gross misregistration, the optimization in the evaluation phase (both for different similarity measures as well as for different subsections) do not need to be as accurate as in the case of the original registration. We use several strategies in order to speed up the computation. The optimization does not need to use the finest level of the multi-resolution pyramid, since the finest level does not lead to larger deviations of the optimum anymore according to our experiments. The threshold that stops the optimization can be chosen larger for the computation of the feature as well.

2.2 Learning Phase

In order to train the learning algorithm, a human observer classifies the performed registrations as successful or failed. We refrained from using several observers, since the inter-observer difference is small in the case of gross misregistration as in our application. The features of the successful and failed registrations are then used as input for a support vector machine (SVM) [14]. Before SVM training, the features are fed into the SVM and first normalized by conversion into Z-scores in order to obtain fewer support vectors during training. A Gaussian radial basis function is used as the kernel of our SVM, which determines the nonlinear mapping from the n-dimensional input feature space into

a higher-dimensional feature space. The SVM is designed to find a max-margin linear classifier separating the two classes (failed registrations versus successful registrations) in the higher-dimensional feature space. The separating hyperplane in the higher-dimensional space is parameterized via support vectors. The SVM classification label for an input feature x can be obtained by considering the sign of

$$f(x) = \sum_{i=1}^{N_s} \alpha_i \gamma_i K(s_i, x) + b \tag{1}$$

where s_i represents the support vectors selected during training and γ_i the corresponding labels (+1/-1), $K(s_i, x)$ denotes the kernel and α_i and b denote the loadings and threshold determined during training. The optimal kernel parameters and the SVM slackness parameter are determined via leave-one-out cross-validation.

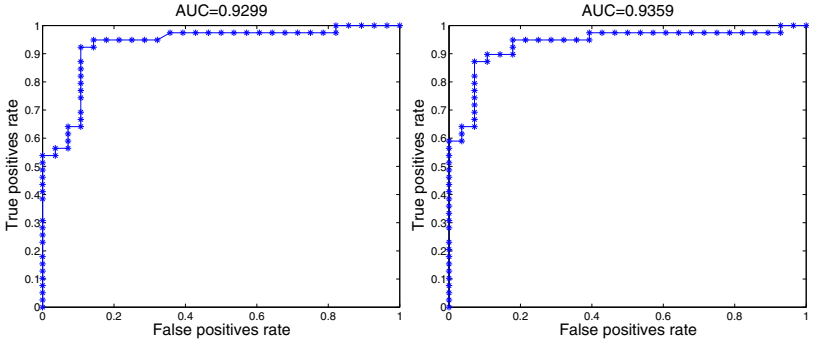


Fig. 4. This figure plots the true positive rate against the false positive rate (equal to 1-false negative rate) for the 5-sim feature set (left) and the corner feature set (right)

3 Experiments and Results

The described method was evaluated for the registration of multiple CT data sets. The test data sets used for the experiments are CT scans showing the head, thorax, abdomen and pelvis, and combinations of these anatomical sections. The volumes overlap completely, partially or not at all. The CT data has different modes of inspiration and different time spans between scans. Since the datasets have a consistent orientation, we perform a rigid registration with the translational component as the free parameters. The optimization algorithm is a standard best-neighbor procedure with normalized mutual information as similarity measure (We still use mutual information for the feature computation, since a deviation by the same measure is a strong indicator for a wrong registration result). The initial parameters of the registration are chosen so that the centers of gravity of both volumes correspond. The optimization starts at the coarsest level of a multiresolution pyramid and proceeds to the next-finer level,

until the improvement in the cost falls below a predetermined threshold. Since we use best-neighbor optimization in the original registration, we use simplex optimization for the evaluation, but we could have used for example a Powell-Brent scheme as well.

From 23 data sets, containing each between 2 and 6 time points with up to 6 CT volumes, we have selected 97 volume pairs and computed their features. We have randomly selected 15 each of the failed and successful registrations and used these as training dataset. The remaining unseen 67 volume pairs have been used for the evaluation of the classifier. The learning has been performed for the features from different similarity measures (5-sim), the features from the subdivisions into corner volumes (corner), the features from the slice-based subdivisions (slices), as well as the combined features from both subdivision methods (combined). The results in the table 1 display the classification results.

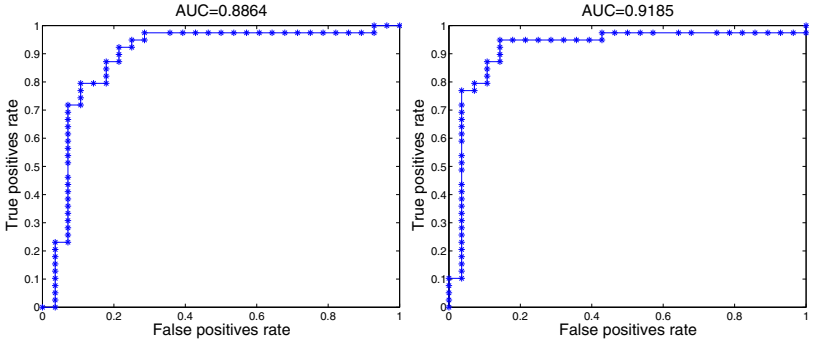


Fig. 5. This figure plots the true positive rate against the false positive rate for the slices feature set (left) and the combined subdivision feature set (right)

Contrary to our expectations, using the 5 different similarity measures has a recognition rate of only 80%, whereas using a subdivision-based approach dividing the moving volume into 8 corners reaches 90%. Adding the feature of the second subdivision approach (slices) improves the result only insignificantly, especially considering the additional computational cost.

Table 1. This table shows the total classification error, the false positive (FP) error, the false negative (FN) error, and the area under the ROC curve (AUC). Note that the AUC $\in [0, 1]$ is invariant to the classifier trade-off between FP and FN errors.

	5-sim	corner	slices	combined
errors	18%	10%	13%	9%
false positives	16%	7%	9%	6%
false negatives	2%	3%	4%	3%
AUC	0.9299	0.9359	8.8864	0.9185

4 Discussion and Future Work

In this work, we have presented a novel method to detect misregistration or suspicious registrations with a 90% success rate for our best feature set. This allows a smoother workflow for offline batch registration and is potentially applicable to many other areas where the misregistration of disparate body parts needs to be detected. The algorithm is straightforward to implement, and we can reduce the cost of the additional registration steps for the evaluation, since the accuracy is less critical for the detection step. The time for the classification is negligible, the learning phase is performed offline and in this case the feature generation (registration) is the expensive step.

With our current set of similarity measures, only CT-CT registration can be evaluated, in future work we would like to explore different sets of similarity measures for different modalities. We expect the subsection-based approach, used with mutual information as in this paper, to work with different modalities, though we have not performed experiments with these modalities so far. Since we have used various anatomical sections (apart from legs), our evaluation is not tuned to a specific region. A more thorough, clinical evaluation with a larger number of datasets as well as with different registration modalities, like MR-MR registration or CT-MR registration is planned for future work as well as exploring more degrees of freedom in the registration algorithm. With deformable registration it might even be possible to distinguish between well-registered and badly-registered regions, and to highlight these.

Another avenue of research is the parallel computation of the different similarity measures. While this is done by Kubias et al. [12] in order to compute a more accurate registration result, the same parallel computation of the similarity measures could be used to detect a divergence between the similarity measures while the computation is still in progress, thus avoiding further costly registration steps.

References

1. Christensen, G.E., Geng, X., Kuhl, J.G., Bruss, J., Grabowski, T.J., Pirwani, I.A., Vannier, M.W., Allen, J.S., Damasio, H.: Introduction to the non-rigid image registration evaluation project (NIREP). In: Pluim, J.P.W., Likar, B., Gerritsen, F.A. (eds.) WBIR 2006. LNCS, vol. 4057, pp. 128–135. Springer, Heidelberg (2006)
2. Hellier, P., Barillot, C., Corouge, I., Gibaud, B., Goualher, G.L., Collins, D.L., Evans, A.C., Malandain, G., Ayache, N.: Retrospective evaluation of inter-subject brain registration. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 258–265. Springer, Heidelberg (2001)
3. Cao, X., Ruan, Q.: A survey on evaluation methods for medical image registration. In: IEEE/ICME International Conference on Complex Medical Engineering, CME 2007, pp. 718–721 (May 2007)
4. Pappas, I.P.I., Stynera, M., Malika, P., Remondab, L., Caversaccioc, M.: Automatic method to assess local CT-MR imaging registration accuracy on images of the head. American Journal of Neuroradiology 26, 137–144 (2005)

5. Kybic, J., Smutek, D.: Image registration accuracy estimation without ground truth using bootstrap. In: Beichel, R.R., Sonka, M. (eds.) CVAMIA 2006. LNCS, vol. 4241, pp. 61–72. Springer, Heidelberg (2006)
6. Kybic, J.: Fast no ground truth image registration accuracy evaluation: Comparison of bootstrap and Hessian approaches. In: ISBI, pp. 792–795. IEEE, Los Alamitos (2008)
7. Kybic, J.: Bootstrap resampling for image registration uncertainty estimation without ground truth. IEEE Transactions on Image Processing 19(1), 64–73 (2010)
8. Wang, H., Feng, D., Yeh, E., Huang, S.: Objective assessment of image registration results using statistical confidence intervals. IEEE Transactions on Nuclear Science 48(1), 106–110 (2001)
9. Schestowitz, R., Twining, C.J., Cootes, T.F., Petrovic, V.S., Taylor, C.J., Crum, W.R.: Assessing the accuracy of non-rigid registration with and without ground truth. In: ISBI, pp. 836–839 (2006)
10. Twining, C.J., Cootes, T., Marsl, S., Petrovic, V., Schestowitz, R., Taylor, C.J.: A unified information-theoretic approach to groupwise non-rigid registration and model building. In: Proceedings of the International Conference on Pattern Recognition (ICPR), pp. 1–14 (2005)
11. Ceylan, C., van der Heide, U.A., Bol, G.H., Lagendijk, J.J.W., Kotte, A.N.T.J.: Assessment of rigid multi-modality image registration consistency using the multiple sub-volume registration (MSR) method. Phys. Med. Biol. 50(10), N101–N108 (2005)
12. Kubias, A., Deinzer, F., Feldmann, T., Paulus, D., Schreiber, B., Brunner, T.: 2D/3D image registration on the GPU. Pattern Recognition and Image Analysis 18(3), 381–389 (2008)
13. Hajnal, J.V., Hill, D.L., Hawkes, D.J.: Medical Image registration. CRC Press, Boca Raton (2001)
14. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines (2001), http://www.csie.ntu.edu.tw/~cjlin/libsvm

Automatic Cortical Gyral Parcellation Using Probabilistic Atlas and Graph Cuts

Gang Li[1], Lei Guo[1], Tuo Zhang[1], Jingxin Nie[1], and Tianming Liu[2]

[1] School of Automation, Northwestern Polytechnical University, Xian, China
[2] Department of Computer Science, The University of Georgia, Athens, GA, USA

Abstract. Automatic parcellation of cortical surfaces into sulci or gyri based regions is of great importance in studying the structure and function of the human brain. This paper presents a novel method for automatic parcellation of cortical surfaces into gyri based regions. The method is composed of two major steps: data-driven gyral patch segmentation and model-driven gyral patch labeling. The gyral patch segmentation is achieved by several steps, including sulcal region segmentation, sulcal basin parcellation, gyral crest segments extraction and gyral patch segmentation. The gyral patch labeling is formulated as an energy minimization problem, in which a cortical probabilistic atlas and the curvature information on surfaces are used to define the energy function. The energy function is efficiently solved by the graph cuts method. A unique feature of the proposed method is that it does not require high dimensional spatial normalization on images or surfaces. The method has been successfully applied to cortical surfaces of 15 young healthy brain MR images. Quantitative and qualitative evaluation results demonstrate the validity of the proposed method.

Keywords: gyral parcellation, gyral basin, probabilistic atlas, graph cuts.

1 Introduction

Automatic parcellation of convoluted cortical surfaces into sulci or gyri based regions is of great importance in studying the structure and function of the human brain. Considerable work has been done for automatic sulcal parcellation of cortical surfaces, since sulci are bounded by gyral crests and can be inferred by using geometric features, such as curvature and sulcal depth. However, gyral parcellation is a much more complicated task due to the fact that most of the gyri are connected together on cortical surfaces and there is no clear boundary between connected gyri [1]. For illustration, Fig 1(b) shows the distinguished gyri and sulci of a cortical surface where red colors indicate sulci and green colors indicate gyri. A few methods have been proposed for cortical gyral parcellation, including atlas based warping methods, where correspondences are established between an expert manually labeled brain atlas and an individual brain image using high dimensional spatial normalization method [2], sulci assisted Voronoi diagram method on cortical surfaces [1], and Bayesian probabilistic labeling method on the spherical coordinate of cortical surfaces [3, 4].

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 438–447, 2010.

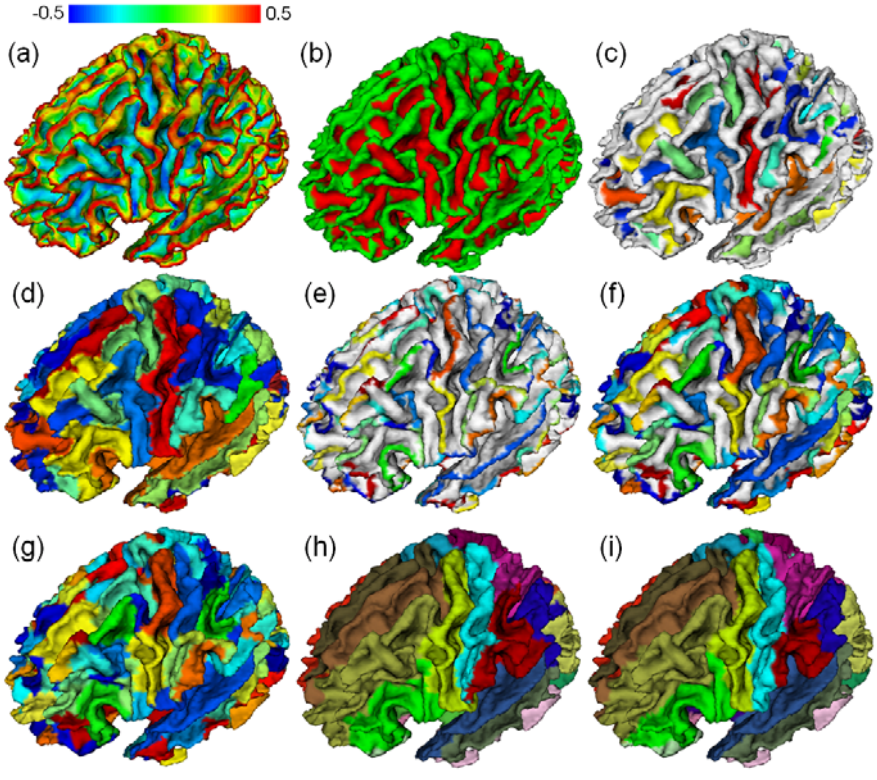


Fig. 1. An example of the proposed method for gyral parcellation on a cortical surface. (a) A cortical surface with the maximum principal curvature map. (b) The distinguished sulcal (red) and gyral (green) regions. (c) The sulcal region segmentation result. (d) The sulcal basin segmentation result. (e) The extracted gyral crest segments. (f) The dilated gyral crest segments. (g) The gyral patch segmentation result. (h) The final gyral patch labeling result. (i) The "ground truth" manually labeled by UCLA LONI experts.

In the paper, a novel method is proposed for parcellation of cortical surfaces into gyral basins [5], which are cortical regions bounded by adjacent sulcal fundi on cortical surfaces. To make the paper more self-contained and easy to understand, before introducing the method, some concepts which will be used in the paper are defined as follows. Sulcal fundi are the curves along the bottoms of sulci on cortical surfaces. Sulcal regions are the buried regions surrounding sulcal space on cortical surfaces. Sulcal basin are the regions bounded by gyral crest lines on cortical surfaces [6]. Gyral crest lines are the curves along the crests of gyri on cortical surfaces. Gyral patches are defined as cortical regions bounded by adjacent sulcal fundi and interrupted at junctions of gyral basins. In general, given a reconstructed cortical surface (herein, the white matter and gray matter interface is adopted), the gyral parcellation method is composed of two major steps: data-driven gyral patch segmentation and model-driven gyral patch labeling. The gyral patch segmentation is achieved by several procedures, including sulcal region segmentation (Fig. 1(c)),

sulcal basin parcellation (Fig. 1(d)), gyral crest segments extraction (Fig. 1(e) and 1(f)) and gyral patch segmentation (Fig. 1(g)), which will be detailed in Section 2. The gyral patch labeling procedure is formulated as an energy minimization problem, in which the spatial prior information from a cortical probabilistic atlas and the curvature information from individual cortical surfaces are used to define the energy function. The energy function is efficiently minimized by the graph cuts method [5], which can guarantee to achieve a strong local minimum for certain energy functions.

Our rationale behind this method is that it combines advantages of both data-driven and model-driven methods. The data-driven method is adopted to effectively partition cortical surfaces into gyral patches. Then the cortical gyral parcellation problem is converted to a gyral patch labeling problem: assigning a gyral basin label for each gyral patch on cortical surfaces. Since a gyral patch can capture more geometric and prior information than a surface point, patch based labeling of cortical surfaces could potentially be more reliable than point based labeling of cortical surfaces. The method is applied to 15 cortical surfaces of young normal subjects and reasonable results have been achieved.

2 Methods

2.1 Gyral Patch Segmentation

Given a reconstructed cortical surface, the proposed gyral parcellation method is composed of two major steps: data-driven gyral patch segmentation and model-driven gyral patch labeling. The gyral patch segmentation is accomplished as follows. First, the cortical surface is segmented into a set of sulcal regions using the hidden Markov random field and Expectation Maximization (HMRF-EM) framework [6] and further parcellated into a set of sulcal basins using the principal direction flow field tracking method [6]. Fig 1(c) and 1(d) show the sulcal region and sulcal basin segmentation results, respectively. Second, the boundaries between parcellated sulcal basins are extracted and all of the boundary vertices linking the same two sulcal basins will be grouped as a gyral crest segment. A morphological dilation, which is constrained in the segmented gyral regions, is further performed on these gyral crest segments. Fig 1(e) and 1(f) show the extracted and dilated gyral crest segments on the cortical surface, respectively. Finally, a principal direction flow field tracking towards gyral crests, based on the flow field inverted from the one used for sulcal parcellation, is performed to partition the cortical surface as gyral patches. In the standard flow field tracking method, every vertex follows the flow field until stopping at a vertex where two successive flow directions pointing to each other. Herein, when performing the flow field tracking, if a vertex on a flow trajectory is in a dilated gyral crest segment, the flow field tracking procedure will be stopped and all the vertices flowing into the same dilated gyral crest segment will be grouped as a gyral patch. This scheme makes the flow field tracking algorithm more robust and less time consuming than the method in [6]. More details of the gyral patch segmentation method is referred to [7]. Fig 1(g) shows the gyral patch segmentation results on the cortical surface.

2.2 Formulation as an Energy Minimization Problem

Since a gyral patch is bounded by adjacent sulcal fundi and interrupted at junctions of gyral basins, a gyral patch typically belongs to only one gyral basin on the cortical surface and each gyral basin is composed of one or multiple gyral patches. Extensive evaluations by experts' visual inspection already confirmed this point (only one example is shown in Fig. 1). Therefore, the cortical gyral parcellation can be considered as a gyral patch labeling problem: assign a label l_p corresponding to a gyral basin to each gyral patch p and meanwhile take into account the contextual information of adjacent gyral patches. This labeling problem can be represented as an energy function, including a data term E_d and a smoothness term E_s:

$$E = E_d + \lambda E_s \tag{1}$$

The data term represents the sum of a set of data costs $D_p(l_p)$ per patch:

$$E_d = \sum_p D_p(l_p) \tag{2}$$

where $D_p(l_p)$ indicates the cost of labeling gyral patch p as gyral basin l_p. In principle, when labeling a gyral patch as the gyral basin to which it belongs, the cost $D_p(l_p)$ should be small. The smoothness term, which represents the sum of the cost of labeling a pair of gyral patches p,q as gyral basins l_p,l_q, is used to impose spatial smoothness. The smoothness term is defined as:

$$E_s = \sum_{\{p,q\}\in N} V_{pq}(l_p,l_q) \tag{3}$$

where N represents the set of neighboring gyral patch pairs. In principle, when labeling two adjacent gyral patches that belong to the same gyral basin as two different gyral basins, $V_{pq}(l_p,l_q)$ should be a large value. Otherwise, when labeling two adjacent gyral patches which inherently belong to two different basins as two different gyral basins, $V_{pq}(l_p,l_q)$ should be a small value. The parameter λ determines the relative contribution between the first data term and the second smoothness term. When λ is set as 0, the data term determines the energy function alone. The optimal labeling of gyral patches is the one which minimizes the above energy function.

2.2.1 Data Term

Because the local geometries of different gyri might be similar, using the information from an individual's cortical surface alone is quite hard to define the data term. Therefore, in our method, the data term is mainly determined by the prior anatomic information of gyri from a cortical probabilistic atlas, which is a powerful tool for segmentation [8]. The probabilistic atlas is constructed from experts manually labeled brain MR images using the method in [9]. Each brain MR image is labeled as a set of gyral structures, and then the labeled images are normalized to the same subject space via affine transform to construct the cortical probabilistic atlas. Herein, the maximum likelihood atlas [9], in which each voxel is labeled as the gyral structure with the

maximum probability, is adopted to define the data term. Using probabilistic atlas instead of the maximum likelihood atlas may further improve the performance of the data term. Note that the subject cortical surface should be transformed into the atlas space. Currently, the affine transformation for normalization of subject images is adopted, and it's found that relatively good results have been achieved. Considering the large inter-subject variations, nonlinear registration might further improve the results, and it will be investigated in the future. However, the gyral patches are segmented on cortical surfaces, while the cortical probabilistic atlas is a volumetric atlas. In order to make use of the atlas, we partition the volumetric image that corresponds to the cortical surface into a set of regions based on the gyral patch segmentation results. As a result, each volumetric region corresponding to a gyral patch is a set of voxels closest to the gyral patch. Fig. 2 shows an example of image partition based on the segmented gyral patches on a cortical surface. Then, we construct a histogram for each volumetric region, corresponding to a gyral patch in the partitioned volumetric image, to represent the likelihood of a gyral patch belonging to the gyral structures in the atlas. Assuming that we have n gyral patches on a cortical surface, the volumetric image will be divided into n corresponding regions, and each region p has a histogram h_p. Then, $h_p(k)$, which indicates the likelihood of gyral patch p belonging to gyral structure k, is calculated as $h_p(k) = (R(p) \bigcap R(k))/R(p)$, where $R(p)$ is the partitioned region p and $R(k)$ is the region of gyral structure k in the maximum likelihood atlas, and $\bigcap$ takes the intersection of the two regions. Then, each h_p is normalized such that $0 \le h_p(k) \le 1$. The larger $h_p(k)$ is, the more likely the gyral patch p belongs to gyral structure k. Therefore, the data term is defined as:

$$E_d = \sum_p D_p(l_p) = \sum_p \sum_k -\ln(h_p(k)) \tag{4}$$

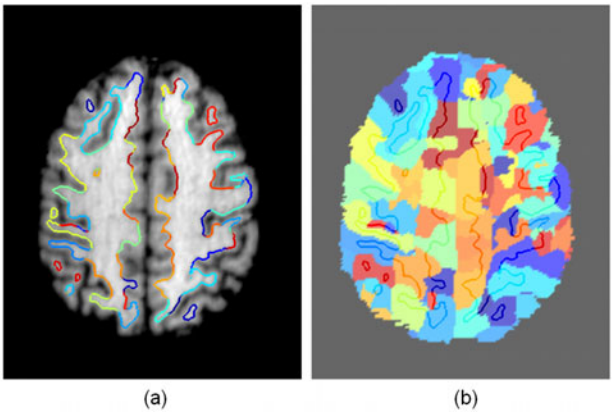


Fig. 2. An example of the volumetric image partition based on gyral patches. (a) The gyral patch segmentation results projected into an image slice. The color curves represent gyral patches. (b) The image partition based on gyral patches. Each region is represented by a color.

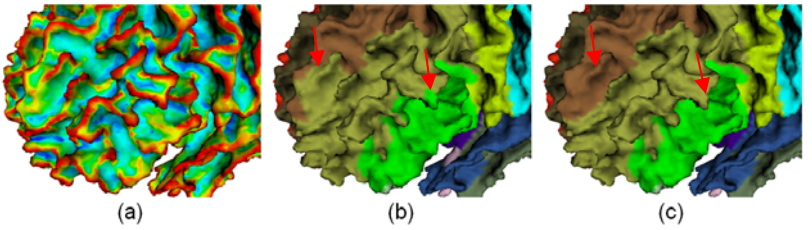


Fig. 3. An example of smoothness term that helps improve the gyral parcellation result. Please see the highlighted areas by red arrows. (a) The maximum principal curvature. (b) The gyral parcellation results with $\lambda = 0$. (c) The gyral parcellation results with $\lambda = 0.3$.

2.2.2 Smoothness Term

To define the smoothness term, we utilize the maximum principal curvatures on cortical surfaces. The maximum principal curvature at each vertex is the principal curvature with largest magnitude in two principal curvatures at the vertex. The gyral crests have large positive maximum principal curvatures, while the sulcal bottoms have large negative maximum principal curvatures [6] as shown in Fig 1(a). Our basic idea is: if two neighboring gyral patches belong to two different gyral basins, they usually meet at sulcal bottoms. Therefore, the average maximum principal curvatures on the boundaries between the two neighboring gyral patches is a large negative value, and the cost of labeling two gyral patches as two different gyri $V_{pq}(l_p, l_q)$ should be a small value. Otherwise, if two neighboring gyral patches belong to the same gyral basin, they will not meet along the sulcal bottom but across the gyral crest, thus the average maximum principal curvatures on the boundaries between the two neighboring gyral patches is close to or larger than zero, and the cost of labeling two gyral patches as two different gyri $V_{pq}(l_p, l_q)$ should be a large value. Therefore, the smoothness term is defined as:

$$V_{pq}(l_p, l_q) = w(p,q)(1 - \delta(|l_p - l_q|)) \tag{5}$$

$$w(p,q) = \exp\Big(\sum_{\{p_i, q_j\} \in \varepsilon} (c(p_i) + c(q_j))/2 \Big/ \sum_{\{p_i, q_j\} \in \varepsilon} 1\Big) \tag{6}$$

where $w(p,q)$ represents the weight between two neighboring gyral patches. δ is defined as: if $\Delta l = 0$, $\delta(\Delta l) = 1$; otherwise, $\delta(\Delta l) = 0$. ε represents the set of all neighboring vertices between two neighboring gyral patches. $c(p_i)$ is the maximum principal curvature at vertex i in gyral patch p. The smoothness term helps improve the gyral labeling results. Fig. 3 shows such an example of comparison between with ($\lambda = 0.3$) and without smoothness term ($\lambda = 0$).

2.3 Graph Cuts for Energy Minimization

Finding a global minimum of the above energy function requires an exhaustive search over the discrete space of all possible labelings. However, several iterative methods can be used to find a strong local minimum of such an energy function, such as the

message passing algorithm [10] and graph cuts [5, 11]. Herein, the graph cuts method is adopted to solve the energy minimization problem. In the graph cuts method, the cortical surface is represented as an undirected weighed graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, where $\mathbf{V}$ is the set of nodes, including all gyral patches and the terminals represented by the set of gyral structures in the maximum likelihood atlas. $\mathbf{E} = \mathbf{E}_N \bigcup \mathbf{E}_T$ is the collection of edges, and $\mathbf{E}_N$ are the edges formed by neighboring gyral patches, called n-links, and $\mathbf{E}_T$ are the edges formed by gyral patches to terminals, called t-links. In above constructed graph, $D_p(\cdot)$ describes the edge weight of t-links, and $V_{p,q}(\cdot)$ describe the edge weight of n-links. A cut $\mathbf{C} \in \mathbf{E}$ is a set of edges by removing which the linked nodes are divided as disjoint sets, in which each node connects to only one terminals corresponding to its gyral basin label. The cost of a cut is the sum of weights on the edge set. Although searching the global optimum of multiple label energy function is NP-hard, graph cuts can guarantee to achieve a strong local minimum efficiently for certain energy functions. More details of graph cuts are referred to [5, 11].

3 Results

To validate the proposed cortical gyral parcellation method, we randomly select 15 subjects which have been manually labeled by UCLA LONI experts from the LONI LPBA atlas [9], which contains 40 labeled subjects. All of the topologically correct and geometrically accurate cortical surfaces used in this section are reconstructed via the method in [12]. For each testing subject, we use the other 39 experts manually labeled subjects to construct a cortical probabilistic atlas, thus leaving out the testing subject. Each probabilistic atlas is reconstructed from the method in [9]. Since manually labeling is performed in volumetric images, the "ground truth" of the gyral parcellation results on cortical surfaces is obtained by assigning the label of the closest voxel in the constructed volumetric maximum likelihood atlas to each vertex. It is noted that the "ground truth" might not be as accurate as directly performing labeling on cortical surfaces. We have tried several different settings for parameter λ, and we found that $\lambda = 0.3$ has achieved relatively good results. Therefore, in all experiments, the parameter λ is set as 0.3. Fig 1(h) and 1(i) show a comparison between an automatically parcellated cortical surface and the "ground truth" on cortical surfaces. Fig. 4 shows the gyral basin segmentation results by our method on the 15 cortical surfaces, where all the gyral basins with identical names are labeled by the same colors. As we can see, almost all the gyral basins have been reasonably parcellated. To quantitatively evaluate the gyral basin segmentation results, we applied the Dice coefficient between the automatically extracted gyral basins and the "ground truth" on cortical surfaces to validate the proposed method. Several major gyri, including precentral gyri, postcentral gyri, superior frontal gyri, middle frontal gyri, inferior frontal gyri, superior parietal gyri, angular gyri, superior temporal gyri, middle temporal gyri, inferior temporal gyri on both left and right hemispheres are used to validate the method. Fig. 5 shows the average Dice coefficients of several major gyral basins on both hemispheres of the 15 subjects, in comparison with the "ground truth" on cortical surfaces. The standard deviation is also shown as error-bar for each gyrus. For example, the average Dice coefficient of left and right precentral

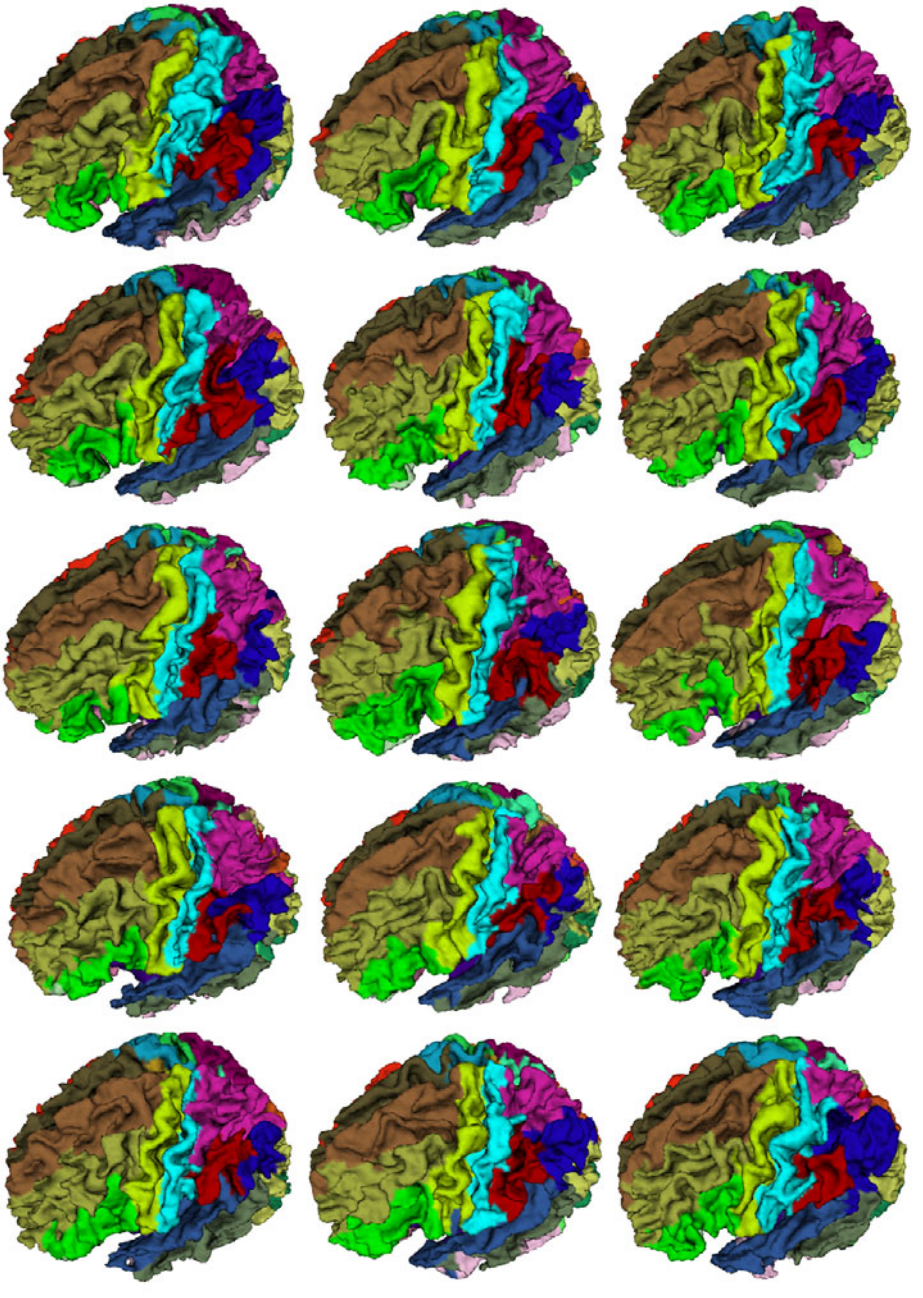


Fig. 4. The gyral basin segmentation results by our method on the 15 cortical surfaces. All the gyral basins with identical names are labeled by the same colors. As we can see, almost all the gyral basins have been visually reasonably parcellated.

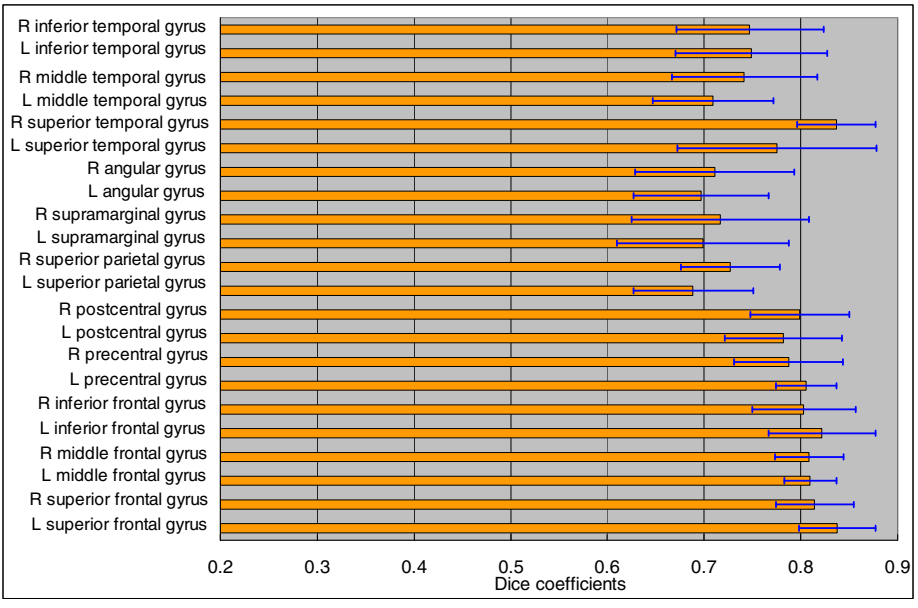


Fig. 5. The average Dice coefficients between the automatically extracted major gyral basins and the "ground truth" by LONI experts on 15 cortical surfaces. The standard deviation is also shown as error-bar for each gyrus.

gyral basins are 0.81 and 0.79 respectively. And the average Dice coefficients of left and right postcentral gyral basins are 0.78 and 0.80 respectively. These results indicate the relatively accurate performance of our gyral basin segmentation method.

4 Conclusion

The paper presents a novel hybrid method for cortical gyral parcellation. In the method, the cortical gyral parcellation is treated as two procedures: data-driven gyral patch segmentation and model-driven gyral patch labeling. The gyral patch labeling is formulated as energy minimization problem. The energy function is defined based on the prior anatomic information from a cortical probabilistic atlas and curvature information. The graph cuts method is adopted to efficiently find a strong local minimum of the energy function. The method has been applied to 15 normal cortical surfaces and validated using the "ground truth" labeled by LONI experts. Our future work includes further improvement and validation of the method using more cortical surfaces.

References

1. Cachia, A., Mangin, J.F., Rivière, D., Papadopoulos-Orfanos, D., Kherif, F., Bloch, I., Régis, J.: A generic framework for the parcellation of the cortical surface into gyri using geodesic Voronoï diagrams. Med. Image Anal. 7(4), 403–416 (2003)
2. Liu, T., Shen, D., Davatzikos, C.: Deformable registration of cortical structures via hybrid volumetric and surface warping. NeuroImage 22(4), 1790–1801 (2004)

3. Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A.M.: Automatically parcellating the human cerebral cortex. Cereb. Cortex 14(1), 11–22 (2004)
4. Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., et al.: An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage 31(3), 968–980 (2006)
5. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. PAMI. 20(12), 1222–1239 (2001)
6. Li, G., Guo, L., Nie, J., Liu, T.: Automatic cortical sulcal parcellation based on surface principal direction flow field tracking. NeuroImage 46(4), 923–937 (2009)
7. Li, G., Guo, L., Li, K., Nie, J., Liu, T.: Gyral parcellation of cortical surfaces via coupled flow field tracking. In: SPIE Medical Imaging, vol. 7623 (2010)
8. Yeo, B.T., Sabuncu, M.R., Desikan, R., Fischl, B., Golland, P.: Effects of registration regularization and atlas sharpness on segmentation accuracy. Med. Image Anal. 12(5), 603–615 (2008)
9. Shattuck, D.W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K.L., Poldrack, R.A., Bilder, R.M., Toga, A.W.: Construction of a 3D probabilistic atlas of human cortical structures. NeuroImage 39(3), 1064–1080 (2008)
10. Murphy, K.P., Weiss, Y., Jordan, M.I.: Loopy belief propagation for approximate inference: An empirical study. In: Proc. of Uncertainty in AI 1999, pp. 467–475 (1999)
11. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts. IEEE Trans. PAMI 26(2), 147–159 (2004)
12. Liu, T., Nie, J., Tarokh, A., Guo, L., Wong, S.T.: Reconstruction of central cortical surface from brain MRI images: method and application. NeruoImage 40(3), 991–1002 (2007)

Hierarchical Fiber Clustering Based on Multi-scale Neuroanatomical Features

Qian Wang[1, 2], Pew-Thian Yap[2], Hongjun Jia[2], Guorong Wu[2], and Dinggang Shen[2]

[1] Department of Computer Science, University of North Carolina at Chapel Hill
qianwang@cs.unc.edu
[2] Department of Radiology and BRIC, University of North Carolina at Chapel Hill
{ptyap,jiahj,grwu,dgshen}@med.unc.edu

Abstract. DTI fiber tractography inspires unprecedented understanding of brain neural connectivity by allowing in vivo probing of the brain white-matter microstructures. However, tractography algorithms often output hundreds of thousands of fibers and thus render the fiber analysis a challenging task. By partitioning a huge number of fibers into dozens of bundles, fiber clustering algorithms make the task of analyzing fiber pathways relatively much easier. However, most contemporary fiber clustering methods rely on fiber geometrical information only, ignoring the more important anatomical aspects of fibers. We propose in this paper a hierarchical atlas-based fiber clustering method which utilizes multi-scale fiber neuroanatomical features to guide the clustering. In particular, for each level of the hierarchical clustering, specific scaled ROIs at the atlas are first diffused along the fiber directions, with the spatial confidence of diffused ROIs gradually decreasing from 1 to 0. For each fiber, a fuzzy associativity vector is defined to keep track of the maximal spatial confidences that the fiber can have over all diffused ROIs, thus giving the anatomical signature of the fiber. Based on the associativity vectors and the ROI covariance matrix, the Mahalanobis distance between two fibers is then calculated for fiber clustering using spectral graph theory. The same procedure is iterated over coarse-to-fine ROI scales, leading to a hierarchical clustering of the fibers. Experimental results indicate that reasonable fiber clustering results can be achieved by the proposed method.

1 Introduction

Diffusion Tensor Imaging (DTI) is well-established for characterizing neural pathways in the brain. It allows microstructural delineation of tissue water diffusion pattern, where water molecules diffuse with more freedom along neurons but not in directions perpendicular to them. Each diffusion tensor captures a part of this diffusion pattern and the tractography algorithms allow a streamline tracing of this pattern along the tensor eigenvectors, resulting in a significant number of fibers. These fibers carry with them abundant brain water diffusivity information and thus provide a unique perspective of neural connectivity as well as the evidence of linkage between brain regions. This has enriched researchers' understanding of the internal working mechanism of the brain, both structurally and functionally.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 448–456, 2010.

A tractography algorithm can usually yield a very huge number of fibers, typically in the order of 10^3-10^6. Such massive amount of fibers often renders subsequent analysis difficult, and makes information provided by the fibers not immediately decipherable. One approach to remedy this is to parcellate tractography results so that the fibers are grouped into the related bundles. For example, fiber clustering can partition the whole tractography results into dozens of bundles, each of which contains fibers behaving similarly in structure as well as in function.

In general, regardless of the clustering method employed, fiber clustering needs the definition of a pairwise similarity/distance measure between two fibers. And most fiber clustering methods up to date rely on fiber geometric features only. In [1] for example, after point-to-point correspondence is detected along two fibers, the ratio measuring the length of the corresponding segment against the overall fiber length is calculated as the similarity measure. The correspondence ratio is highest if the two fibers are identical, and approaches zero if the pairwise correspondence is minimal. In [2], a fiber is regarded as a discrete point set in the Euclidean space, and can be described by the mean of the point set as well as the covariance. Kernel methods are then used to evaluate the fiber similarity based on their individual feature descriptors. In [3][4][5], the generic Hausdorff distance and its variations are applied. The Hausdorff distance in general computes the upper bound of the minimal point-to-point distance between two fibers. Similarity in [6] is estimated by counting the number of voxels through which both fibers are passing, while in [7] the contributions of two fibers are weighted and integrated along the shared pathway. Though no explicit pairwise similarity of fibers is defined, the methods in [8][9][10] adopt a similar strategy when determining the relationship between the fiber under consideration and other fibers (or bundles).

All methods above have in common the shortcoming of not taking into account the neuroanatomical information of fibers, which can be provided by the atlas containing specifications of manual labels. In [11][12][13] for example, information from the atlas is used to help identify detected fiber bundles, though the similarity between fibers is still defined from the geometric perspective, by viewing each fiber as a 3D trajectory only. Recently, some researchers begin introducing the more important anatomical features into fiber clustering. In [14], fibers connecting identical anatomical regions are grouped into a common bundle to reduce the workload of subsequent geometric feature based clustering. However, even though the ROIs are manually delineated by experts, the consistency of the parcellation remains in question due to inevitable bias in labeling. Furthermore, the vector pertaining to the associativity of the fiber with respect to all ROIs is sparse in that a single fiber usually relates to a very limited number of anatomical regions. Great challenges are thus brought into the estimation of pairwise fiber distances based on the anatomical features of fibers.

To solve these problems, we propose a hierarchical atlas-based fiber clustering method which utilizes multi-scale neuroanatomical fiber associativity features. In each level of the hierarchical clustering, a specific scaled ROI from the atlas is diffused along the fiber directions. A fuzzy associativity vector of each fiber is then acquired from the set of diffused ROIs. The Mahalanobis distance between fibers is calculated from their associativity vectors and further employed by spectral clustering. Performed iteratively in a hierarchy of scales, this procedure will eventually lead to a pyramidal clustering of the fibers.

2 Method

The proposed top-down hierarchical fiber clustering scheme relies on multi-scale neuroanatomical fiber features, which will be detailed in Section 2.1. We will define in Section 2.2 the fiber similarity measure in the Mahalanobis distance, and introduce in Section 2.3 the hierarchy of the clustering algorithm.

2.1 Neuroanatomical Features of Fibers

Suppose that there are M ROIs, and the i-th entry ℓ_i in the associativity vector $L_a = (\ell_1, \ell_2, \cdots, \ell_M)$ of fiber a indicates the spatial relationship of the fiber with the i-th ROI. Vector element ℓ_i can be set to 1 if any segment of the fiber lies within the i-th ROI, and set to 0 otherwise. However, the binary formulation as such will result in sparse associativity vectors, pose challenges in estimating fiber distances, and increase the tendency of clustering algorithms to end in local minima.

Here, we propose a fuzzy associativity vector, which takes into account the diffusivity information provided by DTI. Recalling that a typical tractography algorithm traces fibers along the tensor eigenvectors corresponding to the largest eigenvalues, we mimic this process by diffusing each ROI along the fiber directions. Assume that a specific manually delineated ROI covers the domain $\mathcal{L}$ with a constant mass concentration $\phi(x) = \phi_c$. The ROI would then automatically diffuse to the outside of $\mathcal{L}$ due to the imbalance of concentration. We use a generic transport equation to characterize the diffusion of the ROI from $\mathcal{L}$ to the image domain Ω:

$$\nabla_t \phi + \eta(x) f(t, x, \phi, \nabla_x \phi) = g(t, x, \phi), \quad x \in \Omega - \mathcal{L}. \tag{1}$$

In general, the temporal changing rate of the concentration $\nabla_t \phi$ is coupled with the concentration field ϕ and its spatial changing rate $\nabla_x \phi$. The flux term $f(\cdot)$ in Eq. 1 regulates the diffusion velocity by projecting the level-set norms of ϕ onto the space formed by the local tensor $T(x)$:

$$f(t, x, \phi, \nabla_x \phi) = -\left(\frac{\nabla_x^* \phi}{\|\nabla_x^* \phi\|} \cdot T(x) \cdot \frac{\nabla_x \phi}{\|\nabla_x \phi\|}\right) \tag{2}$$

According to Eq. 2, the ROI diffuses rapidly where the norm of its surface is parallel to the eigenvector corresponding to the largest eigenvalue of the local tensor, i.e., coinciding with the fiber pathways. On the contrary, the diffusion would hardly happen if the surface norm of the ROI is nearly perpendicular to the fiber pathways. Meanwhile, the source term $g(\cdot)$ in Eq. 1 enforces the boundary condition $\phi|_{\partial\mathcal{L}} = \phi_c$, preserving the surface ($\partial\mathcal{L}$) of the manually delineated part ($\mathcal{L}$) of the ROI. Further, as in tractography, only voxels whose FA values are sufficiently high (e.g. $\geq$0.2) are considered as possible fiber pathways. We have therefore introduced in Eq. 1 a window function $\eta(\cdot)$ to exclude unrealistic ROI diffusion into low FA areas:

$$\eta(x) = \left(1 + e^{-\alpha(\mathrm{FA}(x) - \beta)}\right)^{-1} \tag{3}$$

The parameter α, which controls the transition steepness of $\eta(\cdot)$, is set to 100 in our case. The center-of-transition cut-off FA value β is set lower than the minimal FA threshold in tractography (e.g. 0.15). The profile of $\eta(\cdot)$ is dependent on the FA value. Fig. 1(b) gives a plot of $\eta(\cdot)$ with $\alpha = 100$ and $\beta = 0.15$ and another with $\alpha = 40$.

An illustration of the ROI diffusion is given in Fig. 1(a), where the underlying tensor principal directions are displayed in green arrows. The zero level-set of the initial ϕ (the dark blue circle) corresponds to the surface of the manually delineated ROI, and is designated as the start time t_0. The anisotropic diffusion is guided by the tensors. The boundary of the ROI surface gradually enlarges and finally deforms to the brown contour at the end time t_1. Also, for each voxel x in the diffused ROI, we define its *spatial confidence* as $(1 - t_\tau)$, based on the time t_τ when the voxel x is traversed by the zero level-set of ϕ. In Fig. 1(a) for example, voxels inside the dark-blue circle (t_0) have a spatial confidence of 1, while the confidence gradually decreases to 0 at the brown contour (t_1).

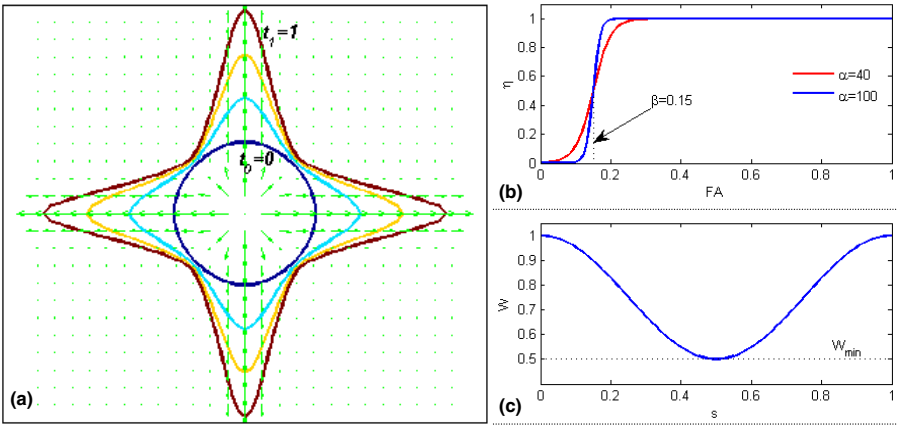


Fig. 1. An example of ROI diffusion is shown in (a), where the ROI surface gradually deforms from the dark-blue circle to the brown contour, guided by the underlying tensor principal directions indicated by green arrows. Panel (b) shows typical profiles of the window function $\eta(\cdot)$ in Eq. 3. Panel (c) provides the importance distribution along the arc-length of the fiber, according to Eq. 4.

Given any fiber a, its *associativity* to the ROI under consideration is given by the maximal spatial confidence of the diffused ROI along the fiber. A fuzzy *associativity vector* L_a can thus be constructed with respect to all M ROIs, and is much less sparse compared to the binary formulation.

The end segments of a fiber are more important than the middle segments in fiber clustering [7]. By normalizing the fiber arc-length s to $[0,1]$, we can specify an *importance* weighting, in the range of $[W_{\min}, 1]$, to different points on the fiber:

$$\begin{aligned} W(s) &= \tfrac{1}{2}(1 - W_{\min})(1 + \cos 2\pi s) + W_{\min} \\ &= \cos^2 \pi s + W_{\min} \cdot \sin^2 \pi s \end{aligned} \tag{4}$$

A typical profile of $W(s)$ is shown in Fig. 1(c), where the minimal importance $W_{\min}$ is set to 0.5. It is worth noting that other forms of importance weighting (e.g., in [7]) are applicable here, since these weighting functions behave similarly by gaining higher importance for the end segments and lower in the middle.

With definition given in Eq. 4, the *importance* value of each ROI with respect to a fiber a can be acquired at the location where the corresponding *associativity* of the fiber a is picked. By considering all M ROIs, an *importance vector* W_a can be produced for each fiber a. Therefore, each fiber is characterized by two vectors – (i) the *associativity vector* carrying neuroanatomical features, and (ii) the *importance vector* weighting the respective neuroanatomical information.

2.2 The Mahalanobis Distance between Fibers

Good fiber clustering result is dependent on a good definition of fiber pairwise distance (or similarity). Pairwise distance between fibers can be computed based on the anatomical feature descriptors. A simple way is to view the associativity vectors as samples in the Euclidean space, where the norm and the inner-product are naturally defined. A better approach, however, is by taking into account the correlation of the ROIs in computing the fiber distance, since ROIs vary significantly in shape and size. By regarding the associativity vector of each fiber as an observation from a multivariate space [15], we can define a covariance matrix which relates the ROIs. Suppose L_a is the associativity vector for the fiber a ($1 \le a \le N$) and $(\cdot)^*$ is the transpose operator, the covariance matrix is:

$$C = \frac{1}{N}\sum_{a=1}^{N} C_a = \frac{1}{N}\sum_{a=1}^{N} \sqrt{L_a \cdot L_a^*} \tag{5}$$

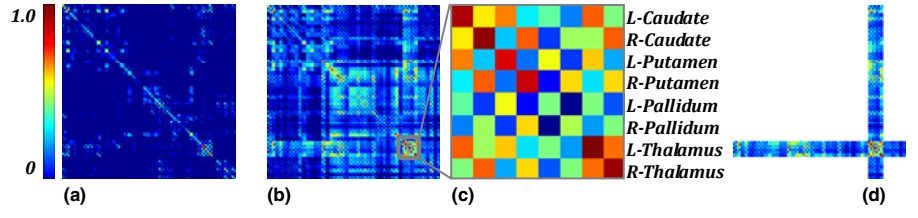


Fig. 2. Example covariance matrices on 90 ROIs are shown, using (a) binary and (b) fuzzy formulation of the associativity measures, respectively. Panel (c) is a close-up of the sub-cortical region in (b), where we can observe higher inside-hemisphere correlation than between-hemisphere. And (d) highlights the significant correlation between sub-cortical areas and other brain regions.

Fig. 2(a-b) give an example of the covariance matrices formed using 90 anatomical ROIs, based on the binary and the fuzzy associativity definitions, respectively. It is obvious that, with fuzzy associativity, the covariance matrix is denser, greatly reducing the sparseness of the associativity vectors. Zooming into the sub-cortical areas in Fig. 2(b), we could observe significant correlation within each hemisphere (Fig. 2(c)), as well as between sub-cortex and other anatomical regions (Fig. 2(d)), as suggested by [16].

Based on the covariance matrix of all ROIs defined in Eq. 5, the Mahalanobis distance between the fiber a and the fiber b can be calculated as follows:

$$d_M(a,b) = \sqrt{(L_a^* - L_b^*) \cdot \mathrm{diag}(W_a \cdot W_b^*)^{\frac{1}{2}} \cdot \mathcal{C}^{-1} \cdot \mathrm{diag}(W_a \cdot W_b^*)^{\frac{1}{2}} \cdot (L_a - L_b)} \quad (6)$$

Then, a corresponding similarity measure, which is to some extent more commonly used in clustering, can be define as $s(a,b) = \exp(-d_M^2(a,b)/2\sigma^2)$, where σ is a predefined value.

2.3 Multi-scale ROI and Hierarchical Clustering Framework

We have adopted a top-down hierarchical mechanism for more principled fiber clustering. In particular, an atlas with three sets of ROIs is registered and aligned to the image space of the DTI dataset. The three multi-scale sets of ROIs consist of 2 ROIs from the two hemispheres, 18 ROIs from various lobes, and finally 90 ROIs delineating different gyri and sulci of the brain [17]. In panel (a) of Fig. 3, we show the three sets of ROIs overlaid on the same typical cortical slice. The different sets of ROIs, therefore, provide information at different scales from the neuroanatomical perspective, giving multi-scale features which can be used to perform a 3-level hierarchical fiber clustering.

In the first level, based on the information given by the 2-ROI set, the tractography results can be easily grouped into 3 classes – fibers in two individual hemispheres and fibers bridging the two hemispheres. In the following two levels of the hierarchy, each class generated in the previous level is further divided into more subclasses, based on the anatomical information of fibers given by the corresponding set of diffused ROIs. As the neuroanatomical features relate to 18 and 90 ROIs in the last two levels, respectively, spectral clustering [18] is recursively called to further cluster fibers based on the similarity matrices of fibers.

Spectral clustering has been widely applied in fiber clustering [2][5], due to its high robustness and efficiency. And in spectral clustering, the number of classes is critical and has to be specified manually. For the convenience to demonstrate the proposed hierarchical fiber clustering method, we simply set the number of classes to 2 in each spectral clustering callback. There are reports in the literature designated to address the issue of optimal class number in spectral clustering [19]. However, the problem indeed exceeds the scope of our work in this paper, though methods to automatically determine the optimal number of classes can be easily integrated into the proposed hierarchical clustering framework.

3 Experimental Results

Fig. 3 shows the results given by the proposed method on approximately 12,000 fibers tracked using ExploreDTI [20] on an adult brain. As can be seen from the top of the figure, the fibers are first categorized at the first level of the hierarchical clustering into 3 classes – fibers in the left/right hemispheres and fibers connecting both hemispheres. In the second and third levels, these classes are further subdivided. We use the same color coding for hemispheric-symmetric fiber bundles. In the third level, the hierarchical clustering strategy yields a total of 12 bundles. And in all three levels, the intra-subject left-right-hemisphere symmetry is well preserved.

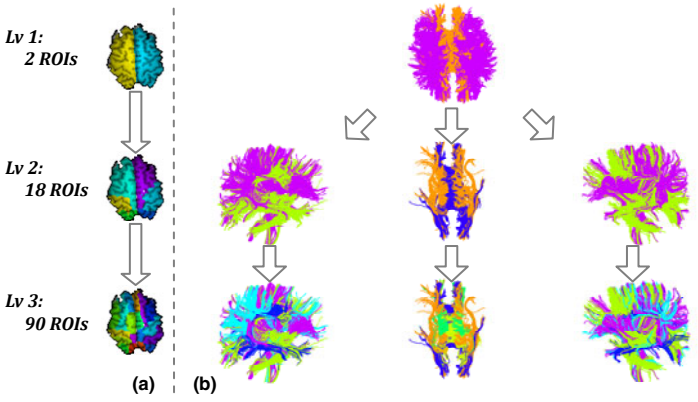


Fig. 3. Hierarchical fiber clustering is performed with the three sets of multi-scale ROIs shown in panel (a). With increasing levels of clustering, more subdivisions of the fibers are obtained, as shown in (b). To highlight the hemispheric symmetry, fiber clustering results in the two hemispheres are displayed in the same color coding.

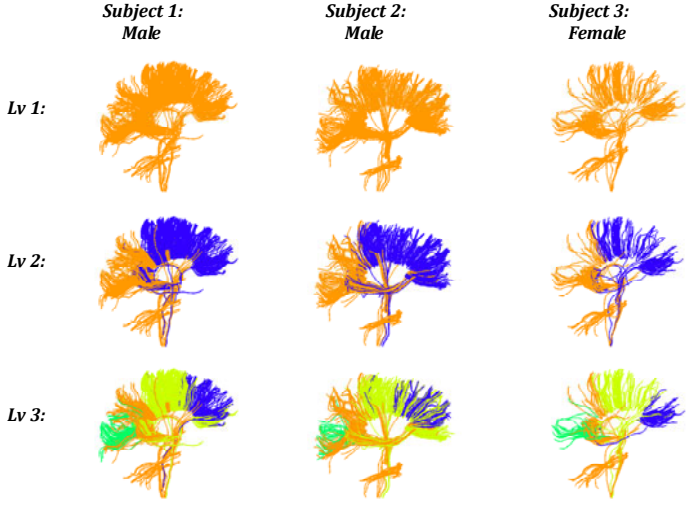


Fig. 4. Consistent fiber clustering across three adult subjects (2 males, 1 female). Here, for better visual inspection, only the fibers connecting the left and the right hemispheres are shown, with same color representing the corresponding bundles in different subjects.

We have further performed the proposed fiber clustering method on three adult brains. For better visualization, only the clustering results of fibers connecting two hemispheres are shown in Fig. 4. The first two subjects are male (~12,000 fibers), while the third one is female and has only ~6,000 fibers yielded by the same tractography algorithm with identical parameters. As we can observe, the inter-subject consistency of clustering is still achieved, due to the use of neuroanatomical information in guiding fiber clustering.

4 Conclusion

We have presented in this paper a hierarchical fiber clustering method. In each level of the hierarchical clustering, our method leverages atlas-based information provided by ROI set of the specific scale. In particular, fuzzy anatomical associativity for each fiber is acquired from the set of diffused ROIs, together with the fiber geometry related importance weighting. The Mahalanobis pairwise distance between fibers is then computed and fed into a spectral clustering algorithm in pyramidal fashion, for performing fiber clustering. Experimental results show that the proposed method obtains good intra-subject symmetry and inter-subject consistency.

References

[1] Ding, Z., Gore, J.C., Anderson, A.W.: Classification and Quantification of Neuronal Fiber Pathways Using Diffusion Tensor MRI. Magnetic Resonance in Medicine 49, 716–721 (2003)

[2] Brun, A., Knutsson, H., Park, H.-J., Shenton, M.E., Westin, C.-F.: Clustering Fiber Traces Using Normalized Cuts. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3216, pp. 368–375. Springer, Heidelberg (2004)

[3] Gerig, G., Gouttard, S., Corouge, I.: Analysis of Brain White Matter Via Fiber Tract Modeling. In: IEEE EMBS 2004, vol. 2, pp. 4421–4424 (2004)

[4] Corouge, I., Gouttard, S., Gerig, G.: Towards a Shape Model of White Matter Fiber Bundles Using Diffusion Tensor MRI. In: ISBI 2004, vol. 1, pp. 344–347 (2004)

[5] O'Donnell, L., Westin, C.-F.: White Matter Tract Clustering and Correspondence in Populations. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 140–147. Springer, Heidelberg (2005)

[6] Jonasson, L., Hagmann, P., Thiran, J.-P., Wedeen, V.J.: Fiber Tracts of High Angular Resolution Diffusion MRI are Easily Segmented with Spectral Clustering. In: ISMRM 2005 (2005)

[7] Klein, J., Bittihn, P., Ledochowitsch, P., Hahn, H.K., Konrad, O., Rexilius, J., Peitgen, H.-O.: Grid-based Spectral Fiber Clustering. In: SPIE Medical Imaging 2007, vol. 6509 (2007)

[8] Maddah, M., Zollei, L., Grimson, W.E.L., Wells, W.M.: Modeling of Anatomical Information in Clustering of White Matter Fiber Trajectories Using Dirichlet Distribution. In: MMBIA 2008 (2008)

[9] Maddah, M., Grimson, W.E.L., Warfield, S.K., Wells, W.M.: A Unified Framework for Clustering and Quantitative Analysis of White Matter Fiber Tracts. Medical Image Analysis 12(2), 191–202 (2008)

[10] Wang, X., Grimson, E., Westin, C.-F.: Tractography Segmentation Using a Hierarchical Dirichlet Processes Mixture Model. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 101–113. Springer, Heidelberg (2009)

[11] Maddah, M., Mewes, A.U.J., Haker, S., Grimson, W.E.L., Warfield, S.K.: Automated Atlas-Based Clustering of White Matter Fiber Tracts from DTMRI. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 188–195. Springer, Heidelberg (2005)

[12] Xia, Y., Turken, A.U., Whitfield-Gabrieli, S.L., Gabrieli, J.D.: Knowledge-Based Classification of Neuronal Fibers in Entire Brain. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 205–212. Springer, Heidelberg (2005)

[13] O'Donnell, L., Westin, C.-F.: Automatic Tractography Segmentation Using a High-Dimensional White Matter Atlas. IEEE Trans. Medical Imaging 26(11), 1562–1575 (2007)
[14] Li, H., Xue, Z., Guo, L., Liu, T., Hunter, J., Wong, S.T.C.: A Hybrid Approach to Automatic Clustering of White Matter Fibers. NeuroImage 49(2), 1249–1258 (2010)
[15] Gong, G., He, Y., Concha, L., Lebel, C., Gross, D.W., Evans, A.C., Beaulieu, C.: Mapping Anatomical Connectivity Patterns of Human Cerebral Cortex Using In Vivo Diffusion Tensor Imaging Tractography. Cerebral Cortex 19(3), 524–536 (2008)
[16] Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C.J., Wedeen, V.J., Sporns, O.: Mapping the Structural Core of Human Cerebral Cortex. PLoS Biology 6(7), 1479–1493 (2008)
[17] Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M.: Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. NeuroImage 15(1), 273–289 (2002)
[18] Shi, J., Malik, J.: Normalized Cuts and Image Segmentation. IEEE Trans. Pattern Analysis and Machine Intelligence 22(8), 888–905 (2000)
[19] Xiang, T., Gong, S.: Spectral Clustering with Eigenvector Selection. Pattern Recognition 41(3), 1012–1029 (2008)
[20] Leemans, A., Jeurissen, B., Sijbers, J., Jones, D.K.: ExploreDTI: A Graphical Toolbox for Processing, Analyzing, and Visualizing diffusion MR Data. In: ISMRM 2009 (2009)

Neural Mass Model Driven Nonlinear EEG Analysis

Xin Fang[1], Zhenghui Hu[1,⋆], Pengcheng Shi[2,⋆]

[1] Department of Optical Engineering Zhejiang University, Hangzhou, 310027, China
zhenghui@zju.edu.cn

[2] B. Thomas Golisano College of Computing and Information Sciences Rochester Institute of Technology Rochester, NY 14623
pengcheng.shi@rit.edu

Abstract. The neural mass models have been widely used for simulating the highly complex Electroencephalogram (EEG) rhythmic activity, when the extrinsic input $p(t)$ passes through the model, similar oscillatory signals are produced. In this paper, we present an empirical exploration to the theoretical prediction of such a model by fitting the actual EEG signal to the Jansen's neural mass model. The results suggest that the model can produce good approximation to the actual EEG signal. The extrinsic input used formerly has a relatively big SD (standard deviation), which may produce unreliable synthetic data, even bias the analysis results. In our study, the mean values of estimated $p(t)$ fall well within the interval for the simulate study recommended by previous reports, but the SD of $p(t)$ is far less than the experience value used before.

Keywords: EEG, Neural mass model, Extrinsic input, Square-root unscented Kalman filter, Nonlinear joint estimation.

1 Introduction

As an integrated biological neural system, which transfers and processes information by way of bioelectricity, the brain comprises local neuronal networks which interconnected with each others by long range pathways. The brain has a hierarchical structure, and one of the most complex parts is the cerebral cortex in the outermost layer of the brain. Electroencephalography (EEG), recording from electrodes placed on the scalp, is the neurophysiologic measurement of the electrical activity of the brain. It derives from the massively summed postsynaptic currents of conglomerated cortical neurons, reflecting the electrical behavior of neuronal interactions. These signals are typically extremely inhomogeneous and non-stationary fluctuations in an irregular and complex manner.

The EEG kinetics could be modeled in a neurophysiological parsimonious scheme of key neuronal mechanism. Neural mass models, one type of nonlinear dynamic systems comprising huge number of neurons, are based upon this knowledge. These models describe populations of neurons interacted through excitation and inhibition and are capable of producing various morphologic EEG-like waveforms and rhythmic activity for the given set of model parameters. There are many attempts involving neural mass models to emulate realistic signals and to evaluate their dynamic properties. Wendling

⋆ Corresponding author.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 457–466, 2010.

et al. used such a model to synthesize activities very similar to those observed in epileptic patients[1]. Niranjan *et al.* adopted a modified method of David and Friston's neural mass model, in order to control epileptic seizures in the model by means of incorporating an internal feedback to maintain synchronous[2]. Roberto *et al.* described a generative model of EEG rhythms based on anatomically constrained coupling of neural mass models[3]. Babajani *et al.* constructed a large-scale biologically realistic neural model for meaningful data fusion of simultaneous EEG and fMRI (functional magnetic resonance imaging) simulated observations[4]. Zavaglia *et al.* ameliorated the neural mass model to simulate EEG power spectral density (PSD) in some regions of interest (ROIs) during simple task[5].

Most existing studies have shown that the output signals produced by the neural mass models are comparable to realistic EEG oscillatory activities, but the chain of the endeavors for analyzing and understanding the bioelectromagnetic signals seems to lack one indispensable part that links the sparing model to actual EEG signal. The neural mass models have been designed to model EEG rhythmic activities, and adjusting parameters in the model can produce various oscillatory signals. Previous studies were concerned over the influence of the variation of model parameters on the oscillatory behavior of the signals generated, but the physiologic significance of some model variables represented respectively were ignored. The extrinsic model input has been arbitrarily modeled as random noise for nonspecific background activity or deterministic function for some specific stimulus activities. However, determining the property of this input signal is extremely important, because, in essence, the neural mass model can be considered as a nonlinear system where a excitatory input signal passing through it will produce the rhythmic activity similar to the EEG signal. In this study, we estimate the extrinsic input by means of the nonlinear joint estimation, the statistic values of the extrinsic input have some difference.

2 Jansen's Neural Mass Model

By the massively synchronous dendritic activity of pyramidal cells, EEG signals are generated to characterize the extremely complex neural networks of brain. The neural mass models of cortical neurons, which offer valuable insight into the generation of the EEG, using a relatively simple nonlinear process to describe the complex neural network. With a small amount of state variables, neural mass models describe the average activity of interaction within and between populations of neurons, these states sum up the behavior of millions of interacting populations.

As with the general models, Jansen's neural mass model also formed by two subsystems[6]. The first subsystem is a wave-to-pulse operator at the soma of neural cells, which can be described by the static sigmoid function[7]:

$$\mathrm{Sigm}_k(v) = \frac{c_k^1 e_0}{1 + exp(r(v_0 - c_k^2 v))} \tag{1}$$

for the kth subpopulation. $c_k^{1,2}$, e_0, r and v_0 are parameters that determine its shape. The second subsystem is a linear pulse-to-wave transform at a synaptic level. The former transforms the average pulse density of action potentials coming to the population

of neurons into an average postsynaptic potential (PSP) which can either be excitatory or inhibitory, this is assumed to be instantaneous. The later transforms the average membrane potential of a population of neurons into an average pulse density of action potentials fired by the neurons, which depends on synaptic kinetics and models the average postsynaptic response as a linear convolution of incoming spike rate. Between two subsystems, there are also some connectivity constants $C_1, \cdots, C_4$ which account for the number of synapses that established between two neurons populations.

These relations are formulated as:

$$\begin{cases} \ddot{x}_0(t) = \frac{H_e}{\tau_e}\mathrm{Sigm}[x_1(t) - x_2(t)] - \frac{2}{\tau_e}\dot{x}_0(t) - \frac{1}{\tau_e^2}x_0(t) \\ \ddot{x}_1(t) = \frac{H_e}{\tau_e}\{p(t) + C_2\mathrm{Sigm}[C_1 x_0(t)]\} - \frac{2}{\tau_e}\dot{x}_1(t) - \frac{1}{\tau_e^2}x_1(t) \\ \ddot{x}_2(t) = \frac{H_i}{\tau_i}\{C_4\mathrm{Sigm}[C_3 x_0(t)]\} - \frac{2}{\tau_i}\dot{x}_2(t) - \frac{1}{\tau_i^2}x_2(t) \end{cases} \tag{2}$$

where x_0, x_1 and x_2 are respectively the output of the three PSP blocks, $p(t)$ is the excitatory input that represents an average firing rate, which can be random (accounting for a nonspecific background activity) or deterministic, accounting for some specific activities in other cortical units. The typical value of model parameters for simulation experiments are shown in Table **??** [8]. There are three second-order time derivatives variables x_0, x_1 and x_2 in this system, the general way to reveal the relationship between these parameters is by means of the *Laplace transform*, but the *Laplace transform* is powerless when the system is nonlinear and time-variant.

The state space model, which was generally applied in control engineering, represents a physical system as n first order coupled differential equations, i.e., for an nth order differential education would be transformed to n first order ordinary differential equations(ODEs). Thus we rewrite the system mentioned above as a set of six first order ODEs by introducing three new variables $\dot{x}_0(t) = x_3(t)$, $\dot{x}_1(t) = x_4(t)$ and $\dot{x}_2(t) = x_5(t)$.

In Jansen's neural mass model, a cortical area is composed of three different populations of neurons (excitatory pyramidal cells, excitatory stellate cells, and inhibitory interneurons). For the two basic reasons: (1) the excitatory and inhibitory synaptic inputs on the pyramidal neurons lead to generate the potential field; (2) pyramidal neurons in the form of a palisade oriented perpendicular to the cortex, thus the EEG signal is considered as the membrane potential of pyramidal neurons. For the sake of simplicity, we do not take into account the effect of the amplifiers that are used to record the EEG nor the distributions produced by the different tissue layers between the neural mass and the recording electrode [6,7]. As mentioned above, x_1 and x_2 are the result of excitatory and inhibitory synapses acting on the pyramidal neurons, hence, the EEG observation is taken as: $y(t) = x_1(t) - x_2(t)$. Different from *Laplace transforms* which deals the system with the frequency-domain approach, time-domain approach was introduced to model and analyze systems in the state space model. At the same time, the limitations of 'linear system' and 'zero initial conditions' was broken.

For the actual EEG signal acquisition equipment, the faint electrical activity (In fact, EEG activity reflects the summation of the synchronous activity of thousands or millions of neurons which are radial to the scalp.) is measured by putting electrodes on different positions of the scalp. EEG is caused by correlated post-synaptic potentials of cortical neurons (As voltage fields decrease with the fourth power of the radius, it is more difficult to detect action potentials from the deep sources than the cortex.), that

Table 1. Physiological interpretation and standard values of model parameters

Parameter	Physiological interpretation	Standard value
$H_{e,i}$	Average synaptic gain	$H_e = 3.25mV$ $H_i = 22mV$
$\tau_{e,i}$	Membrane average time constant and dendritic tree average time delays	$\tau_e = 10ms$ $\tau_i = 20ms$
$c_1^{1,2}$	Average number of synaptic contacts in the excitatory feedback loop	$c_1^1 = c$, $c_1^2 = 0.8c, c = 135$
$c_2^{1,2}$	Average number of synaptic contacts in the inhibitory feedback loop	$c_2^1 = 0.25c$ $c_2^2 = 0.25c$
$c_3^{1,2}$		$c_3^1 = c_3^2 = 1$
ν_0, e_0, r	Parameters of the nonlinear sigmoid function	$\nu_0 = 6mV, e_0 = 5s^{-1}$ $r = 0.56\ mV^{-1}$

means there are some minor differences to generate the action potentials between different correlated pyramidal neurons, from a macro perspective, EEG shows oscillations at a variety of frequencies in the same channel while they are similar between different channels. In our model, we consider that all of the activities of the different neuronal populations are modulated by the same specific stimulus-dependent input $p(t)$ which represents the basal stochastic activity, thus, the $y(t)$ of different channels would be integrated into a vector $\mathbf{y}(t)$ which contains n elements, where n related to the numbers of channels:

$$\mathbf{y}(t) = \mathbf{C}[x_1(t) - x_2(t)] \tag{3}$$

where $\mathbf{C}$, which is also a vector contains n elements, relates to the weights of linear relationship between voltage signal detected from the scalp and post-synaptic potentials of pyramidal neurons. There should be some small differences between the elements of $\mathbf{C}$, for the sake of simplicity, here we assume that the value of all elements equal to constant 1.

We model the EEG process as a deterministic function, which capture the underline pattern, plus the error term. The processes $\mathbf{x}$ and observations $\mathbf{y}$ are composed of the fixed effects and a stochastic part. Thus the corresponding state space model of the system as mentioned above would be a two-layer model. The internal layer is a nonlinear stochastic difference equation:

$$\dot{\mathbf{x}}_k = f(\mathbf{x}_k, p_k) + \nu \qquad \nu \sim N(0, \mathbf{Q}) \tag{4}$$

The external layer is assumed to the measurement equation:

$$\mathbf{y}_k = h(\mathbf{x}_k) + \omega \qquad \omega \sim N(0, \mathbf{R}) \tag{5}$$

where f and h are process and observation equations, respectively, $\mathbf{x}_k = [x_0, x_1, \cdots, x_5]^T$ is the state of the system, the extrinsic inputs p_k represents system input, ν is the noise process caused by disturbances and modeling errors, $\mathbf{y}_k$ is the observation vector, and ω is measurement noise.

3 Nonlinear Joint Estimation

As a recursive data processing algorithm, the Kalman filter (KF) is the most general approach for analyzing and making inference for such a state space model. The KF is an optimal estimator for linear systems that are driven by *Gaussian* noise, and are observed through linear means which may also incorporate with *Gaussian* errors. It consists of essentially two stages: the prediction stage, which uses the system model to predict the state posterior density (pdf) forward from one measurement time to the next, and the update stage, which uses the latest measurement to modify the prediction pdf. But the inherent flaws of KF is that it can only solve the discrete-data linear filtering problem.

For nonlinear state-space estimation, we use a square-root unscented Kalman filter (SR-UKF), which is a derivative-free alternative to the extend Kalman filter in nonlinear case, to maintain the nonlinearities present in the biophysical models. SR-UKF propagates variables mean and covariance through the *unscented transformation* (UT), and possesses high accuracy and robustness for nonlinear model estimations. The framework of the state-estimation algorithm by using SR-UKF is given as follows[9]:

Step 1. By calculating the matrix square-root of the state covariance once through a *Cholesky* factorization, SR-UKF is initialized:

$$\hat{x}_0 = \mathbb{E}[x_0] \qquad \mathbf{S}_0 = chol\{E[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T]\} \tag{6}$$

Step 2. For $k \in \{1, \cdots, \infty\}$, Sigma points is calculated as:

$$\mathcal{X}_{k-1} = [\hat{x}_{k-1} \quad \hat{x}_{k-1} + \eta \mathbf{S}_k \quad \hat{x}_{k-1} - \eta \mathbf{S}_k] \tag{7}$$

Step 3. Time updates:
(1) The converted set is given by transforming each point through the process equation:

$$\mathcal{X}_{k|k-1} = \mathbf{F}[\mathcal{X}_{k-1}, \mathbf{u}_{k-1}] \tag{8}$$

(2) The predicted mean is calculated as:

$$\hat{x}_k^- = \sum_{i=0}^{2L} W_i^{(m)} \mathcal{X}_{i,k|k-1} \tag{9}$$

(3) With a *QR* decomposition, the *time-update* of the *Cholesky* factor $\mathbf{S}_k^-$ is calculated:

$$\mathbf{S}_k^- = qr\{[\sqrt{W_1^{(c)}}(\mathcal{X}_{1:2L,k|k-1} - \hat{x}_k^-) \quad \sqrt{\mathbf{R}^v}]\} \tag{10}$$

(4) It is essential to do *Cholesky* update because the $\mathbf{W}_0^{(c)}$ may be negative:

$$\mathbf{S}_k^- = cholupdate\{\mathbf{S}_k^-, \mathcal{X}_{0,k} - \hat{x}_k^-, \mathbf{W}_0^{(c)}\} \tag{11}$$

(5) Instantiate each of the prediction points through the observation equation:

$$\mathcal{Y}_{k|k-1} = \mathbf{H}[\mathcal{X}_{k|k-1}] \tag{12}$$

(6) The predicted observation is calculated as:

$$\hat{y}_k^- = \sum_{i=0}^{2L} W_i^{(m)} \mathcal{Y}_{i,k|k-1} \tag{13}$$

Step 4. Measurement update:
(1) The innovation *Cholesky* factor is computed as:

$$\mathbf{S}_{\tilde{y}_k} = qr\{[\sqrt{W_1^{(c)}}(\mathcal{Y}_{1:2L,k|k-1} - \hat{y}_k^-) \qquad \sqrt{\mathbf{R}_k^n}]\} \tag{14}$$

(2) Cholesky update because the $\mathbf{W}_0^{(c)}$ may be negative:

$$\mathbf{S}_k^- = cholupdate\{\mathbf{S}_{\tilde{y}_k}, \mathcal{Y}_{0,k} - \hat{y}_k, \mathbf{W}_0^{(c)}\} \tag{15}$$

(3) The cross-covariance matrix of x and y is determined by:

$$\mathbf{P}_{x_k y_k} = \sum_{i=0}^{2L} W_i^{(c)}[\mathcal{X}_{i,k|k-1} - \hat{x}_k^-][\mathcal{Y}_{i,k|k-1} - \hat{y}_k^-]^T \tag{16}$$

(4) The Kalman gain matrix can be found according to

$$\mathcal{K}_k = (\mathbf{P}_{x_k y_k} / \mathbf{S}_{\tilde{y}_k}^T) / \mathbf{S}_{\tilde{y}_k} \tag{17}$$

(5) The update mean is calculated as:

$$\hat{x}_k = \hat{x}_k^- + \mathcal{K}_k(y_k - \hat{y}_k^-) \tag{18}$$

(6) The posterior measurement update of the *Cholesky* factor of the state covariance is calculated (the downdate vectors are the columns of $\mathbf{U}$) by:

$$\mathbf{U} = \mathcal{K}_k \mathbf{S}_{\tilde{y}_k} \tag{19}$$

$$\mathbf{S}_k = cholupdate\{\mathbf{S}_k^-, \mathbf{U}, \mathbf{-1}\} \tag{20}$$

where $\mathbf{R^v}$=process noise *cov*., $\mathbf{R^n}$=measurement noise *cov*..

In our case, both the system state $\mathbf{x}_k$ and the input variable $p(t)$ for the dynamic system must be simultaneously estimated from only the observed noisy signal $\mathbf{y}_k$.the joint state vector is $\tilde{\mathbf{x}} = [p, x_0, x_1, x_2, x_3, x_4, x_5]^T$ here.

The initial extrinsic input $p(t)$ was set as $p_0 = 200$ in the range reported by Jansen *et al.* [6]. Thus, the initial condition was set as $\tilde{\mathbf{x}}_0 = [200, 0, 0, 0, 0, 0, 0]^T$.

4 Result and Discussion

In this experiment, 10-channel EEG time series were recorded for 18 experiment subjects, and all with eyes closed. For each subject, the measurement was performed for an average recording time of 130 seconds, at a frequency of 250 Hz[10]. The Jansen's

Table 2. The statistic values of estimated $p(t)$. 18 subjects were selected to collect EEG signal at resting status (eyes closed and relax), the statistic mean values were around 200 while the standard deviation values were small (with the order of magnitude from 0.01 - 1).

Experiment	Statistic Values of $p(t)$		Experiment	Statistic Values of $p(t)$	
Subjects	Mean	S.D.	Subjects	Mean	S.D.
Sub_{01}	199.79	0.070	Sub_{10}	198.32	1.224
Sub_{02}	198.26	1.110	Sub_{11}	200.16	0.034
Sub_{03}	194.20	0.780	Sub_{12}	202.46	0.344
Sub_{04}	198.36	0.239	Sub_{13}	195.46	2.408
Sub_{05}	200.07	0.014	Sub_{14}	197.08	2.702
Sub_{06}	200.02	0.042	Sub_{15}	199.90	0.071
Sub_{07}	199.84	0.250	Sub_{16}	199.61	0.269
Sub_{08}	201.03	0.245	Sub_{17}	199.63	0.431
Sub_{09}	200.13	0.043	Sub_{18}	200.00	0.015

neural mass model was fitted to each recording using the SR-UKF algorithm described above. Before the iteration process, ODE45, which implemented the *Runge-Kutta-Fehlberg* method with a variable time step for efficient computation, was used to translate the continuous equations of the model into discrete time series model.

The traditional Jansen's neural mass model, for example, the Jansen's double-column model, which was explored using two coupled model columns, considered that the external input should be different for each column. Although the proportions of cortical cells (e.g., the visual cortex and the prefrontal cortex) differ from each others, different model parameter values could be used to express those differences. In our study, all outputs collected from different columns were considered to be stimulated by the same external input, i.e., with a 10-channel EEG collection equipment, we need a vector contains 10 elements to describe the different relationship between the respective output and external input. In fact, the basic neuronal architecture of the cortex is similar, so the difference is slight. For the consideration of simplicity, we assume that the value of all vector elements equal to 1.

Although Jansen's neural mass model is a simplified nonlinear dynamic system, analogous oscillatory behavior compared with the real EEG signal was produced. In Figure 2 we show an episode of the real EEG time series (red) of channel *o1* and the estimated rhythmic activity (blue) , the corresponding power spectral density distribution reflects the amount of similarity with respective frequency components. The results suggest that EEG oscillatory activity can be described by a neurophysiological neural mass model. At the same time, it also provides some support for the hypothesis that the spontaneous EEG activity is generated by some of the same neural structures. We collected the EEG signal at resting status, three types frequency band of EEG were observed obviously: (1) δ (0.1-3 Hz), which represents deep sleep and unconscious; (2) θ (4-8 Hz), which denotes the situations of drowsiness, arousal and idling; (3) α (8-12 Hz), which indicates the status of relaxed and reflecting, and this three bands EEG signal could be clearly seen at the bottom of Figure 2.

Table 2 shows the statistic values of the estimated extrinsic input $p(t)$. There are several previous reports proposed that the mean values of extrinsic input $\bar{p} \approx 220$. Jansen *et*

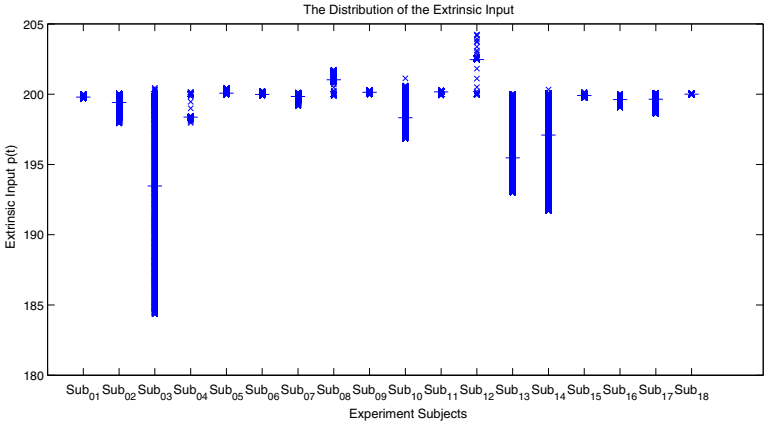


Fig. 1. The distribution of mean values of the extrinsic input $p(t)$. $'\times'$ denotes the value of $p(t)$ estimated by fitting the actual EEG signal to model, $'-'$ denotes the mean value of all estimated values.

al. estimated the model parameters by using the $p(t)$ with a big standard deviation $\sigma_p = 57.7$. Although the signal produced by the estimated models resembled the signals generated by the reference models, significant difference was found between the estimated parameters and the actual values[11]. Valdes *et al.* used the nonlinear Kalman filter and the maximum likelihood procedure to estimate the variance of $p(t)$ by fitting and validating theoretical models for actual EEG data, much smaller standard deviation of $p(t)$ ($\sigma_p \approx 4.5$) was estimated[12]. Recently, Ponten *et al.* analyzed the relationship between structural and functional connectivity, the extrinsic input they used was Gaussian white noise with standard deviation of 0.1 potentials per second[13].

In our study, for all 18 experiment subjects, $\bar{p} \approx 200$, is well within the value interval. But the standard deviation shows a noticeable difference from the previous researches. Actually, excepting 4 subjects (Sub_2, Sub_{10}, Sub_{13}, Sub_{14}), the standard deviation of $p(t)$ for all other subjects is less than 0.8 (even most of them less than 0.3). So there are very small fluctuations of the extrinsic input we estimated, and it could be observed intuitively in Figure 1.

It is recognized that the values of model parameters play a more important role in simulating EEG rhythmic activity than extrinsic input. For models of multiple areas, we need adjust the relative proportion of each population in the cortical area. The parameters w^n, n=[1, ..., N], bounded between 0 and 1, was used to implement this. David have detected that oscillations were produced even without extrinsic input ($\sigma_p = 0$) when $w = 0$ or $w = 1$[7]. But a too big p variance still may produce unreliable synthetic data, thus could bias the analysis results. For the bigger p standard deviation, since the model input $p(t)$ represents the activity of multi-cortical columns, a high variance value may be contributed to recruit a larger number of coupled neuron for complex process that causes complex input signal dynamics. In other words, difference variance may indicate the activity extent in specific neuron area.

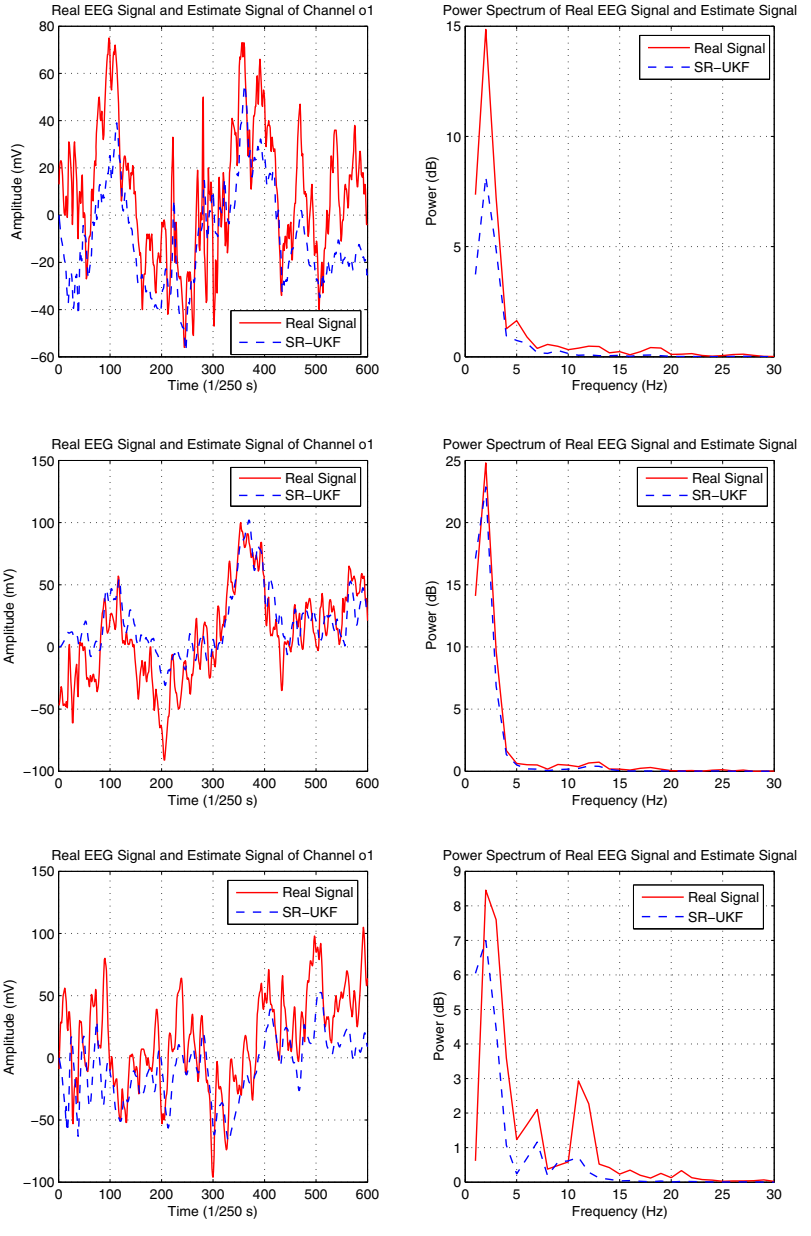


Fig. 2. Left: the real EEG time series (red, solid line) and estimated signal (blue, dotted line) based on neural mass model. Right: the corresponding power spectral density distribution of real signals (red, solid line) and estimated signal (blue, dotted line). Top: channel $o1$ placed on Sub_{02}; Middle: channel $o1$ placed on Sub_{03}; Bottom: channel $o1$ placed on Sub_{06}.

In conclusion, by fitting and validating the Jansen's neural mass model for real EEG data, extrinsic input $p(t)$ is estimated. Unlike the $p(t)$ used to produce the similar EEG signal, which suffered big variance, the $p(t)$ estimated here have a much smaller standard deviation, this will be a basis for us to improve the traditional neural mass model.

Acknowledgments. This work has been supported by the National Basic Research Program of China (No: 2010CB732500), the National Natural Science Foundation of China (No: 30800250), Doctoral Fund of Ministry of Education of China (No: 200803351022), Zhejiang Provincial Natural Science Foundation of China (No: Y2080281), and Zhejiang Provincial Qianjiang Talent Plan (No: 2009R10042).

References

1. Wendling, F., Bellanger, J.J., Bartolomei, F., Chauvel, P.: Relevance of nonlinear lumped-parameter models in the analysis of depth-eeg epileptic signals. Biological Cybernetics 83, 367–378 (2000)
2. Niranjan, C., Sabesan, S., Tsakalis, K., Iasemidis, L.: Controlling epileptic seizures in a neural mass model. J. Comb. Optim. 17, 98–116 (2009)
3. Sotero, R.C., Barreto, N.J.T., Medina, Y.I., Carbonell, F., Jimenez, J.C.: Realistically coupled neural mass models can generate EEG rhythms. Neural Computation 19, 478–512 (2007)
4. Babajani, A., Soltanian-Zadeh, H.: Integrated MEG/EEG and fMRI model based on neural masses. IEEE Transactions on Biomedical Engineering 53(7), 1794–1801 (2006)
5. Zavaglia, M., Astolfi, L., Babiloni, F., Ursino, M.: A neural mass model for the simulation of cortical activity estimated from high resolution EEG during cognitive or motor tasks. Journal of Neuroscience Methods 157, 317–329 (2006)
6. Jansen, B.H., Rit, V.G.: Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. Biological Cybernetics 73, 357–366 (1995)
7. David, O., Friston, K.J.: A neural mass model for MEG/EEG: coupling and neuronal dynamics. NeuroImage 20, 1743–1755 (2003)
8. David, O., Cosmelli, D., Friston, K.J.: Evaluation of different measures of functional connectivity using a neural mass model. NeuroImage 21, 659–673 (2004)
9. van der Merwe, R., Wan, E.A.: The square-root unscented Kalman filter for state and parameter-estimation. In: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 6, pp. 3461–3464 (2001)
10. Hu, Z.H., Shi, P.C.: Regularity and Complexity of Human Electroencephalogram Dynamics: Applications to Diagnosis of Alzheimers Disease. In: IEEE International Conference on Pattern Recognition (ICPR), vol. 3, pp. 245–248 (2006)
11. Jansen, B.H., Kavaipatti, A.B., Markusson, O.: Evoked potential enhancement using a neurophysiologically-based model. Method Inform. Med. 40, 338–345 (2001)
12. Valdes, P.A., Jimenez, J.C., Riera, J., Biscay, R., Ozaki, T.: Nonlinear EEG analysis based on a nerual mass model. Biological Cybernetics 81, 415–424 (1999)
13. Ponten, S.C., Daffertshofer, A., Hillebrand, A., Stam, C.J.: The relationship between structural and functional connectivity: Graph theoretical analysis of an EEG neural mass model. NeuroImage (2009), doi:10.1016/j.neuroimage.2009.10.049

Modeling the Dermoscopic Structure Pigment Network Using a Clinically Inspired Feature Set

Maryam Sadeghi[1,2,3], Majid Razmara[1], Paul Wighton[1,2,3], Tim K. Lee[1,2,3], and M. Stella Atkins[1]

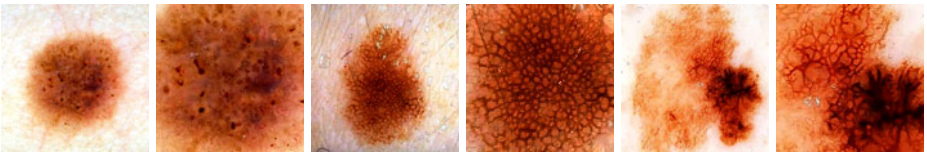
[1] School of Computing Science, Simon Fraser University, Canada
msa68@sfu.ca
[2] Department of Dermatology and Skin Science, University of British Columbia and Vancouver Coastal Health Research Institute, Canada
[3] Cancer Control Research Program and Cancer Imaging Department, BC Cancer Research Centre, Canada

Abstract. We present a method to detect and classify the dermoscopic structure pigment network which may indicate early melanoma in skin lesions. We locate the network as darker areas constituting a mesh, as well as lighter areas representing the 'holes' which the mesh surrounds. After identifying the lines and holes, 69 features inspired by the clinical definition are derived and used to classify the network into one of two classes: *Typical* or *Atypical*. We validate our method over a large, inclusive, real-world dataset consisting of 436 images and achieve an accuracy of 82% discriminating between three classes (*Absent*, *Typical* or *Atypical*) and an accuracy of 93% discriminating between two classes (*Absent* or *Present*).

1 Introduction

Melanoma, a cancerous lesion in the pigment-bearing basal layers of the epidermis, is the most deadly form of skin cancer, yet treatable via excision if detected early. The cure rate for early-stage melanoma is nearly 100%. A recent study [1] has concluded that dermoscopy increases the early detection of melanoma, only if the practitioner is sufficiently trained. In fact, dermoscopy *decreases* accuracy if training is insufficient. There is, therefore, a demand to develop computer-aided diagnostic systems to facilitate the early detection of melanoma. This paper follows a relatively new trend in clinical dermatology: to identify specific 'dermoscopic structures' in the lesions such as the pigment network which is then used to arrive at a diagnosis [2]. A pigment network can be classified as either *Typical* or *Atypical*, where a working definition of a typical pigment network (TPN) is "a light-to-dark-brown network with small, uniformly spaced network holes and thin network lines distributed more or less regularly throughout the lesion and usually thinning out at the periphery" [3]. For an atypical pigment network (APN) we use the working definition "a black, brown or gray network with irregular holes and thick lines" [3]. The goal is to automatically classify a given image to one of three classes: *Absent*, *Typical*, or *Atypical*. Figure 1 exemplifies these 3 classes.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 467–474, 2010.



(a) (b) (c) (d) (e) (f)

Fig. 1. The 3 classes of the dermoscopic structure pigment network: a-b) *Absent*; c-d) *Typical*; and e-f) *Atypical*. b),d),f) are magnifications of a),c),e) respectively.

We use these definitions to subdivide the structure into the darker mesh of the pigment network (which we refer to as the 'net') and the lighter colored areas the net surrounds (which we refer to as the 'holes'). After identifying these substructures we use the definitions above to derive a several structural, geometric, chromatic and textual features suitable for classification. The result is a robust, reliable, automated method for identifying and classifying the structure pigment network.

2 Previous Work

The automated detection of pigment network has received some recent attention [4,5,6,7,8,9,10]. Fleming et al. [4] report techniques for extracting and visualizing pigment network via morphological operators. Fischer et al. [5] use local histogram equalization and gray level morphological operations to enhance the pigment network. Anantha et al. [6] propose two algorithms for detecting pigment networks in skin lesions. They are most successful when employing a weighted average of two Laws' energy masks whose weights are determined empirically and report a classification accuracy of approximately 80%. Betta et al. [7] begin by taking the difference of an image and its response to a median filter. This difference image is thresholded to create a binary mask which undergoes a morphological closing operation to remove any local discontinuities. This mask is then combined with a mask created from a high-pass filter applied in the Fourier domain to exclude any slowly modulating frequencies. Results are reported graphically, but appear to achieve a sensitivity of 50% with a specificity of 100%. Di Leo et. al. [8] extend this method and compute features over the 'holes' of the pigment network. A decision tree is learnt in order to classify future images and an accuracy of 71.9% is achieved. Shrestha et. al. [9] begin with a set of 106 images where the location of the APN has been manually segmented. If no APN is present, then the location of the most 'irregular texture' is manually selected. They then compute several texture metrics over these areas (energy, entropy, etc.) and employ various classifiers to label unseen images. They report accuracies of approximately 95%. Recently, Sadeghi et al. [10] presented an approach whereby 'holes' of the network are detected using a graph-based loop search on the result of an edge detection algorithm. They then create a graph

based on the distance of these 'holes' which they use to propose a new spatial feature called 'density ratio' for detecting a pigment network in a given image.

Although these studies have certainly made significant contributions, there has yet to be a comprehensive analysis of pigment network detection on a large number of dermoscopic images under 'real-world' conditions. All work to date has either: 1) not reported quantitative validation [4,5]; 2) validated against a small ($n < 100$) number of images[7]; 3) only considered or reported results for the 2-class problem (e.g. *Absent*/*Present* rather than *Absent*/*Typical*/*Atypical*) [6,7,8,9,10]; 4) not explicitly identified the location of the network [6]; or 5) has made use of unrealistic exclusion criteria and other manual interventions [9].

This paper presents an effective method for pigment network segmentation and classification, which is validated on a large ($n = 436$) 'real-world' dataset.

3 Method

An overview of our method for the identification and classification of pigment network is given in Fig. 2. After pre-processing, Sadeghi et al.'s [10] 'hole detector' is employed to generate a 'hole mask' indicating the pixels belonging to the holes of the pigment network. Next, a 'net mask' is created, indicating the pixels belonging to the net of the pigment network. We then use these masks to compute a varity of features including structural (which characterizes shape), geometric (which characterizes distribution and uniformity), chromatic and textural features. These features are fed into a classifier to classify unseen images.

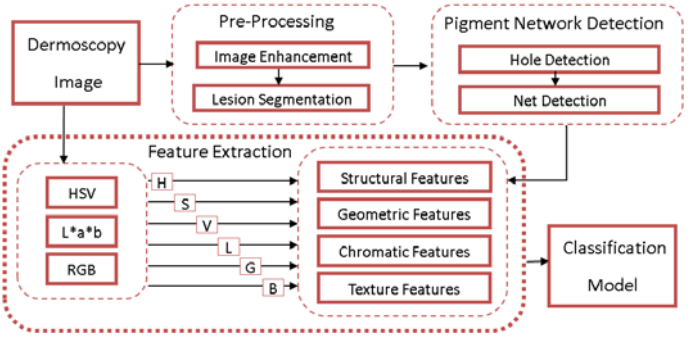


Fig. 2. Overview of construction of our classification model

3.1 Pre-Processing

Image Enhancement: The image is first enhanced so that the pigment network is more visible using manner similar to [5]. A contrast-limited adaptive histogram equalization is used [11]. Images are then further sharpened by subtracting a blurred version of the image from itself.

Lesion Segmentation: Next, in order to prevent unnecessary analysis of the pixels belonging to the skin, lesions are segmented using Wighton et. al.'s method [12] which employs supervised learning and the random walker algorithm. The output of the segmentation is a 'lesion mask' which indicates which pixels belong to the lesion (as opposed to the surrounding healthy skin).

3.2 Pigment Network Detection

Hole Detection: To find the holes of the pigment network, we used Sadeghi et al.'s method [10] which employs a Laplacian of Gaussian edge detector to create a graph-based structure which is used to identify holes of the pigment network. They demonstrate that *globules* (another dermoscopic structure which is difficult to discriminate from *pigment netowrk*) are separable from the holes of the pigment network according to the difference between the average intensity of inner pixels and the average intensity of the border [10]. The result of this process is a 'hole mask' which indicates which pixels belong to the holes of the pigment network.

Net Detection: In order to identify the net of a pigment network, we apply the Laplacian of Gaussian (LoG) filter to the green channel of the image. The LoG filter identifies high frequency components of an image and therefore makes an ideal net detector. The major issue with applying this operator is that its response is strongly dependent on the relationship between the frequency of the structures and the size of the Gaussian kernel used. We used $\sigma = 0.15$, which is an appropriate value for images of the two atlases used in our experiment [13,14], however it can be tuned for a given imageset according to scale and magnification. In our experiment, we observed that the average thickness of the pigment network is proportional to the average size of holes of the network. We therefore set the size of the LoG window size to half of the average hole size in the image. The average window size over all images of our data set is 11 pixels. We then threshold the filter response, resulting in a 'net mask' which indicates which pixels belong to the net of the the pigment network. Furthermore, we skeletonize this mask, resulting in a 'skeleton mask'. Figure 3 illustrates the net extraction process.

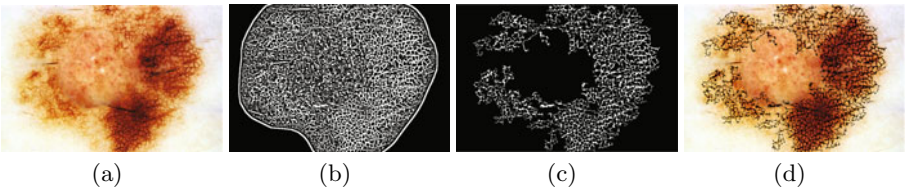


Fig. 3. Net detection. a) A dermoscopic image, b) response of the LoG filter, c) the resulting 'net mask', and d) the extracted net of the pigment network overlaid on the original image

3.3 Feature Extraction

Based on our working definitions of TPN and APN, we use the resulting masks of sections 3.2 and 3.1 (namely the lesion, hole, net and skeleton masks) to propose a set of features capable of discriminating between the 3 classes (*Absent*, *Typical* and *Atypical*). We propose a set of structural (shape), geometric (spatial) chromatic and textural features.

Structural Features (20): Diagnostically important characteristics of an *Atypical* network include the thickness and variation in thickness of net as well as the size and variation in size of network holes.

For each spatially disjoint section of the net mask, we compute its size (number of pixels in the net mask) and length (number of pixels in the skeleton mask). Our features are then the mean, standard deviation and ratio ($mean/std$) of the sizes and lengths of net sections. Thickness is also computed by measuring the distance from each pixel in the net mask to the closest pixel in the skeleton mask. The mean, standard deviation and ratio of thickness as well as a 6-bin thickness histogram are also included as features.

For each spatially disjoint section of the hole mask, we compute the size (number of pixels) and include as features the mean, standard deviation and ratio ($mean/std$) of hole size as well as the total number of holes.

We also include the ratio of the network size (number of pixels in the net and hole masks) to the lesion size (number of pixels in the segmentation mask).

Geometric Features (2): Clinically, there is an emphasis on the 'uniformity' of the network in order to differentiate between TPN and APN. Sadeghi et. al. [10] have proposed the feature 'density ratio' of holes which, while useful in discriminating between the absence and presence of a pigment network, does not reliably discriminate between the TPN and APN. We include 'density ratio' of holes as a feature, as well as a new feature, 'hole irregularity'. This feature is computed by constructing a graph as in [10] where 2 holes are connected if they are less than 3 times the average diameter of the holes. 'Hole irregularity' is then the number of edges in this graph.

Chromatic Features (37): Color also plays a crucial role in clinical diagnosis. We therefore convert the image to HSV colourspace and compute features over each channel as well as the original green channel of the image. In each channel, for the hole, net and lesion masks respectively we compute the mean, standard deviation and ratio ($mean/std$) of the intensity values. Additionally, we also propose a new chromatic feature called the 'atypicality measure' which is the sum of the intensity values over the green channel of the pixels in the net mask.

Textural Features (10): We use five of the classical statistical texture measures of Haralick et al. [15]: entropy, energy, contrast, correlation and homogeneity which are derived from a grey level co-occurrence matrix (GLCM). The GLCM is a tabulation of how often different combinations of pixel brightness values (gray levels) occur in a specific pixel pairing of an image. We construct

2 GLCMs and extract the 5 texture metrics from each. The first GLCM is constructed over the entire lesion (using the pixels in the lesion mask) and the second is constructed over the pigment network (using the pixels in the net and hole masks).

3.4 Classification

Finally, these 69 features are fed into a classifier so that new images can be classified. We employ the WEKA's [16] implementation SimpleLogistic which uses a powerful boosting algorithm LogitBoost. Boosting is a method for combining the performance of many features to produce a powerful classifier. SimpleLogistic fits logistic models by applying LogitBoost with simple regression functions as base learners.

4 Evaluation and Results

We applied the method described above to a set of dermoscopic images taken from two atlases of dermoscopy [13,14]. In [13] each image is labeled as *Absent*, *Typical* or *Atypical*, representing the presence and regularity of the dermoscopic structure pigment network. However the images in [14] have been labeled by 40 experts, each one assigning a label of either *Absent*, *Typical* or *Atypical* to each image. Overall labels for these images are generated by majority voting. In total, our dataset consists of 436 images (161 *Absent*, 154 *Typical*, 121 *Atypical*). We compute results for both the 3-class (*Absent*, *Typical* or *Atypical*) and 2-class problems (*Absent*, *Present*). Ten-fold cross validation was used to generate all results. For comparison, the feature set described in [8] was also implemented and results over our imagesets computed. Table 1 summarizes these results.

Table 1. Comparing accuracy, precision, recall and f-measure of our proposed features with Di Leo et al.'s features using the same set of 436 images. The last three rows summarize the results from previous work on different image sets.

Absent-Typical-Atypical Classification					
	Precision	Recall	F-measure	Accuracy	N
Absent	0.905	0.950	0.927	-	161
Typical	0.787	0.792	0.790	-	154
Atypical	0.750	0.694	0.721	-	121
Weighted Avg	**0.820**	**0.823**	**0.821**	**0.823**	436
Di Leo et al. [8]	0.709	0.711	0.709	0.719	436
Absent-Present Classification					
Absent	0.893	0.932	0.912	-	161
Present	0.959	0.935	0.947	-	275
Weighted Avg	**0.935**	**0.933**	**0.934**	**0.933**	436
Di Leo et al. [8]	0.875	0.876	0.875	0.876	436
Absent-Present Classification on Different Image Sets					
Anantha et al. [6]	-	-	-	0.80	155
Betta et al. [7]	-	-	-	0.66	30
Sadeghi et al. [10]	-	-	-	0.926	500

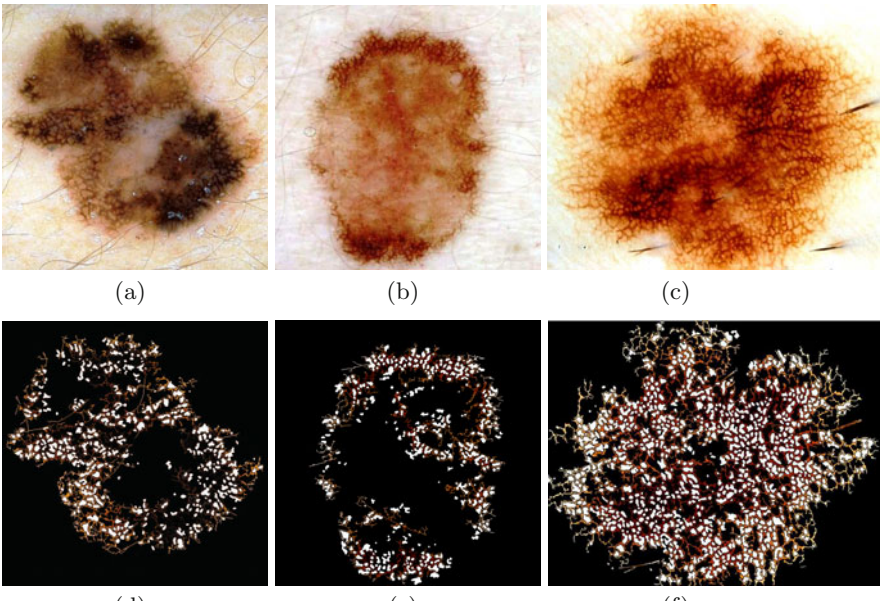


Fig. 4. Three images of the image set: the top row shows the original images of a APN, APN, and TPN. The bottom row shows their corresponding pigment networks.

Comparing these results with the results generated by the others using different datasets may be tenuous, nevertheless these reported values are also summarized in Table 1. As can be seen, this work outperforms other previous work on the 2-class problem and is the only one to report quantitative results of the 3-class problem. Additionally, qualitative results of detecting pigment network 'net' and 'holes' is illustrated in Figure 4.

5 Conclusion and Future Work

We have described techniques to identify the sub-structures of the dermoscopic structure pigment network. Furthermore, we have proposed and validated a set of clinically motivated features over these sub-structures suitable for classification. Our feature set has proven to be extremely robust, outperforming previous work on a more inclusive 'real-world' dataset consisting of 436 images, which is the largest validation to date on the 3-class problem.

This feature set can be used to aid in the automation of clinical dermoscopic algorithms [2]. Future work will be motivated by the notion of a fully automatic, robust automation of clinical dermoscopic algorithms and will therefore focus on reliably identifying and classifying other dermoscopic structures in a similar fashion.

Acknowledgements

This work was funded by the Canadian Natural Sciences and Engineering Council (NSERC), CIHR-Skin Research Training Centre (SRTC) and a grant from the Canadian Health Research Project (CHRP).

References

1. Kittler, H., Pehamberger, H., Wolff, K., Binder, M.: Diagnostic accuracy of dermoscopy. The Lancet Oncology 3(3), 159–165 (2002)
2. Johr, R.: Dermoscopy: alternative melanocytic algorithm–the ABCD rule of dermatoscopy, Menzies scoring method, and 7-point checklist. Clinics in Dermatology 20(3), 240–247 (2002)
3. Argenziano, G., Soyer, H., et al.: Dermoscopy of pigmented skin lesions: results of a consensus meeting via the Internet. Journal of the American Academy of Dermatology 48(5), 679–693 (2003)
4. Fleming, M., Steger, C., et al.: Techniques for a structural analysis of dermatoscopic imagery. Computerized Medical Imaging and Graphics 22(5), 375–389 (1998)
5. Fischer, S., Guillod, J., et al.: Analysis of skin lesions with pigmented networks. In: Proc. Int. Conf. Image Processing, Citeseer, pp. 323–326 (1996)
6. Anantha, M., Moss, R., Stoecker, W.: Detection of pigment network in dermatoscopy images using texture analysis. Computerized Medical Imaging and Graphics 28(5), 225–234 (2004)
7. Betta, G., Di Leo, G., et al.: Dermoscopic image-analysis system: estimation of atypical pigment network and atypical vascular pattern. In: MeMea 2006, pp. 63–67. IEEE Computer Society, Los Alamitos (2006)
8. Di Leo, G., Liguori, C., Paolillo, A., Sommella, P.: An improved procedure for the automatic detection of dermoscopic structures in digital ELM images of skin lesions. In: IEEE VECIMS, pp. 190–194 (2008)
9. Shrestha, B., Bishop, J., et al.: Detection of atypical texture features in early malignant melanoma. Skin Research and Technology 16(1), 60–65 (2010)
10. Sadeghi, M., Razmara, M., Ester, M., Lee, T., Atkins, M.: Graph-based Pigment Network Detection in Skin Images. In: Proc. of SPIE, vol. 7623 (2010)
11. Zuiderveld, K.: Contrast limited adaptive histogram equalization. In: Graphics gems IV, pp. 474–485. Academic Press Professional, Inc., London (1994)
12. Wighton, P., Sadeghi, M., Lee, T., Atkins, M.: A fully automatic random walker segmentation for skin lesions in a supervised setting. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 1108–1115. Springer, Heidelberg (2009)
13. Argenziano, G., Soyer, H., et al.: Interactive atlas of dermoscopy. EDRA-Medical Publishing and New Media, Milan (2000)
14. Soyer, H., Argenziano, G., et al.: Dermoscopy of pigmented skin lesions. In: An Atlas Based on the Consensus Net Meeting on Dermoscopy 2000, Edra, Milan (2001)
15. Haralick, R., Dinstein, I., Shanmugam, K.: Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics 3(6), 610–621 (1973)
16. Goebel, M.: A survey of data mining and knowledge discovery software tools. ACM SIGKDD Explorations Newsletter 1(1), 20–33 (1999)

An Application Driven Comparison of Several Feature Extraction Algorithms in Bronchoscope Tracking During Navigated Bronchoscopy

Xióngbiāo Luó[1], Marco Feuerstein[1,2], Tobias Reichl[2], Takayuki Kitasaka[3], and Kensaku Mori[4,1]

[1]Graduate School of Information Science, Nagoya University, Japan
[2]Computer Aided Medical Procedures, Technische Universität München, Germany
[3]Faculty of Information Science, Aichi Institute of Technology, Japan
[4]Information and Communications Headquarters, Nagoya University, Japan

Abstract. This paper compares Kanade-Lucas-Tomasi (KLT), speeded up robust feature (SURF), and scale invariant feature transformation (SIFT) features applied to bronchoscope tracking. In our study, we first use KLT, SURF, or SIFT features and epipolar constraints to obtaininter-frame translation (up to scale) and orientation displacements and Kalman filtering to recover an estimate for the magnitude of the motion (scale factor determination), and then multiply inter-frame motion parameters onto the previous pose of the bronchoscope camera to achieve the predicted pose, which is used to initialize intensity-based image registration to refine the current pose of the bronchoscope camera. We evaluate the KLT-, SURF-, and SIFT-based bronchoscope camera motion tracking methods on patient datasets. According to experimental results, we may conclude that SIFT features are more robust than KLT and SURF features at predicting the bronchoscope motion, and all methods for predicting the bronchoscope camera motion show a significant performance boost compared to sole intensity-based image registration without an additional position sensor.

Keywords: Bronchoscope Tracking, Camera Motion Estimation, KLT, SURF, SIFT, Image Registration, Navigated Bronchoscopy.

1 Introduction

In minimally invasive diagnosis and surgery of lung and bronchus cancer, a physician usually performs transbronchial needle aspiration (TBNA) to obtain tissue samples to assess suspicious tumors as well as to treat or remove precancerous tissue. However, it is difficult to properly navigate the biopsy needle to the region of interest (ROI) for sampling tissue inside the airway tree, because the TBNA procedure is usually guided by conventional bronchoscopy, which only provides 2-D information (bronchoscopic video images), and because of the complexity of the structure of the bronchial tree. Recently, bronchoscopic navigation systems have been developed to guide the TBNA procedure by fusing pre-operative and

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 475–484, 2010.

intra-operative information such as 3-D multi-detector computed-tomography (CT) image data and real-time bronchoscopic video. This helps a physician to properly localize the biopsy needle during navigated bronchoscopy.

For navigated bronchoscopy the exact pose of the bronchoscope camera must be tracked inside the airway tree. Unfortunately it is really challenging to accurately track the position and orientation of the bronchoscope camera inside the patient's airway tree in real time during bronchoscopic navigation. So far, two main approaches (or their combination) for bronchoscope tracking have been proposed in the literature: (a) sensor-based and (b) vision-based tracking. The former uses an electromagnetic (EM) tracking system (e.g., the superDimension navigation system [12]) to locate an electromagnetic sensor that is usually fastened at the bronchoscope tip to directly measure the bronchoscope camera position and orientation. The latter analyzes the bronchoscopic video images obtained from the bronchoscope camera to continuously track the bronchoscope tip on the basis of image registration methods [10,6]. This is a widely discussed topic in the field of bronchoscope tracking and also the topic of our paper.

Usually, vision-based methods use image registration techniques to align a real bronchoscope camera pose to a virtual camera pose generated by placing a virtual camera inside the 3-D CT data. However, a major drawback is that image registration techniques heavily depend on characteristic information of bronchial trees (e.g., bifurcations or folds), so they can fail easily to track the bronchoscope camera in the case of the shortage of such information [5]. Feature-based bronchoscope motion estimation is a promising means for dealing with this problem during bronchoscope tracking [10,4]. Without any characteristic information, other texture feature information of real bronchoscopic video frames can be extracted and used to compensate the performance of image registration.

Basically, a feature-based approach for motion estimation and recovery first needs to extract features from camera images, which can be utilized to compute the relative camera motion, for example by epipolar geometry (up to scale). Currently, two well-known methods for extracting features are the SURF and SIFT algorithm [2,8]. Both return distinctive features from keypoints that are invariant to image scale and rotation. Also, the KLT tracker first detects good features by calculating the minimum eigenvalue of each 2×2 gradient matrix and selects features to be tracked using an optimization (e.g. Newton-Raphson) method for minimizing the difference between two feature windows from two consecutive images [11].

However, little work can be found that evaluates the effectiveness of these different feature extraction algorithms that are used for bronchoscope tracking during bronchoscopic navigation. This study utilizes these feature-based camera motion tracking methods to improve the performance of image registration-based bronchoscope tracking. We use the KTL, SURF, and SIFT features to estimate inter-frame pose displacements (up to scale) on the basis of epipolar constraints and Kalman filtering to get position estimates before performing image registration. We compare and evaluate the respective performances of KLT, SURF, and SIFT features used for bronchoscope tracking.

2 Method

Feature-based camera motion estimation algorithms are widely used in the field of structure from motion (SFM) or stereo vision. These approaches basically consist of two main steps: (1) feature extraction and (2) feature tracking. The first step usually characterizes some points or regions in each video image as interest features that carry motion information among video images. Sequentially, inter-frame motion parameters (up to scale) can be estimated in the second step by recognizing corresponding features between consecutive video frames. In our work, we detect interest features for each real bronchoscopic (RB) video image using a KLT-, SURF-, or SIFT-based method, respectively. We address the difficulty of determining the magnitude of motion (here referred to as scale factor) by Kalman filtering during feature-based motion estimation.

Our proposed bronchoscope tracking method has two major stages: rough camera motion estimation and intensity-based image registration. Figure 1 displays a flow-process diagram of our tracking method. First, KLT, SURF, or SIFT features are respectively detected from the current bronchoscopic video image and feature correspondences are identified in the previous frame. During epipolar geometry analysis, inter-frame camera motion up to scale is predicted on the basis of these feature correspondences. Kalman filtering is then applied to estimate the uncertain scale factor, or in other words, the magnitude of the relative motion. Finally, after combining the estimates of epipolar geometry analysis and Kalman filtering to a full Euclidean transformation matrix that moves the

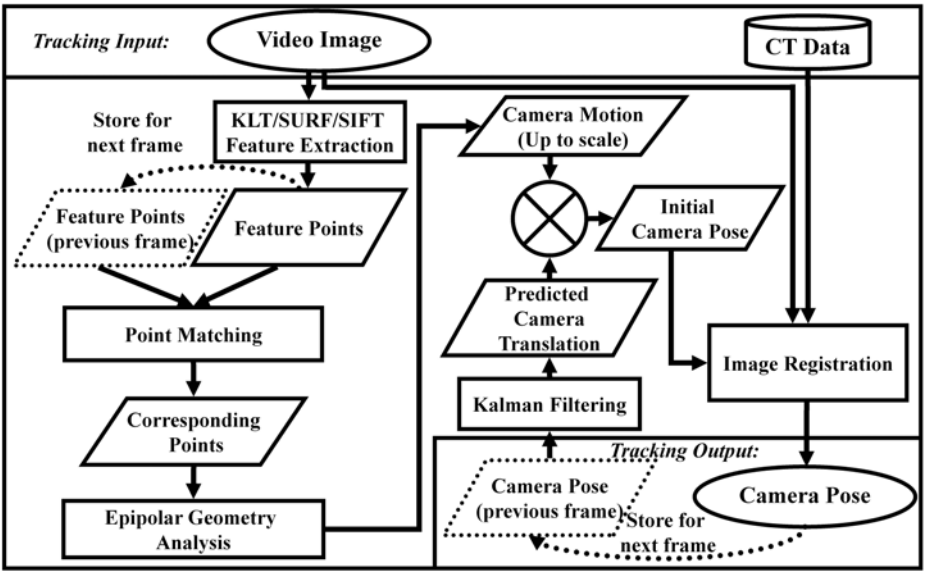


Fig. 1. Processing flowchart of our motion tracking method

camera from the previous to the current pose, we can perform image registration initialized with this matrix.

Specifically, the feature-based bronchoscope camera motion tracking process is performed by the following five steps:

[Step 1] Feature detection. We extract 2-D feature points by using the KLT, SURF, or SIFT algorithm [1,2,8]. The KLT tracker is sometimes referred to as corner detector while the other two approaches, which are considered as scale invariant feature detectors, try to find characteristic blob-like structures in an image independent of its actual size. The SURF or SIFT detector can be constructed using a scale space representation of an original image at different resolutions. After detecting feature points, SIFT usually describes each feature point using a 128-dimensional vector while SURF does so with a 64-dimensional vector. All these vectors include the local gradient direction and magnitude information in a certain square neighborhood centered at the feature point. More details about these feature detection algorithms can be found in the original publications [1,2,8]. We note that the normal SURF algorithm is implemented by doubling the initial image resolution in our case, and hence we can obtain good performance, as shown in the work of Bauer et al. [1].

[Step 2] Feature correspondences. After feature point detection from bronchoscopic video sequences, we must determine feature correspondences that can be used to find the relative motion relation between two successive RB images. The KLT method extracts adequate feature points of an RB image and uses normalized cross correlation (NCC) to track (or match) them. However, for SURF or SIFT feature points, we recognize corresponding 2-D point pairs using the third matching strategy from the work of Mikolajczyk and Schmid [9]. Additionally, a simple outlier detection mechanism was performed on the basis of the standard deviation of the distances between corresponding points to remove unsuitable point pairs.

[Step 3] Epipolar geometry analysis. Inter-frame motion parameters $\Delta\tilde{\mathbf{Q}}^{(i)}$ between the (i-1)-th and (i)-th RB image contain a translation unit vector $\Delta\tilde{\boldsymbol{t}}^{(i)}$ and rotation matrix $\Delta\tilde{\mathbf{R}}^{(i)}$ that can be predicted with epipolar geometry analysis by solving the following equations sequentially:

$$\mathbf{E}^T \Delta\tilde{\boldsymbol{t}}^{(i)} = \mathbf{0}, \tag{1}$$

$$\Delta\tilde{\mathbf{R}}^{(i)}\mathbf{E}^T = \left[\Delta\tilde{\boldsymbol{t}}^{(i)}\right]_{\times}^{T} \tag{2}$$

where $\mathbf{E}$ is the essential matrix described epipolar constraints [7] that our corresponding points must satisfy. It needs to be clarified that the essential matrix $\mathbf{E}$ involves an arbitrary scale factor. Hence the absolute translation vector $\Delta\hat{\boldsymbol{t}}^{(i)}$ depends on an arbitrary scale factor $\tilde{\alpha}^{(i)} = |\Delta\tilde{\boldsymbol{t}}^{(i)}|$ that depicts the real magnitude of the translational motion. An effective method to predict this scale that is based on Kalman filtering is proposed in the next step.

[Step 4] Kalman filtering-based scale factor estimation. Kalman filtering is widely developed in the community for target position tracking on the basis of a state-space model [3]. In our work, Kalman-based motion filtering is employed to determine the magnitude of the bronchoscope translational motion. Basically, the scale factor $\hat{\alpha}^{(i)}$ can be determined by

$$\hat{\alpha}^{(i)} = |\Delta\hat{\boldsymbol{t}}^{(i)}| \;\; = \;\; |\hat{\boldsymbol{t}}^{(i)} - \hat{\boldsymbol{t}}^{(i-1)}|, \tag{3}$$

where the camera absolute translation vector $\hat{\boldsymbol{t}}^{(i-1)}$ and $\hat{\boldsymbol{t}}^{(i)}$ are calculated by Kalman filtering.

We can now retrieve the absolute translation vector $\Delta\tilde{\boldsymbol{t}}_*^{(i)}$ between frames $(i-1)$ and i from the unit transaltion vector $\Delta\tilde{\boldsymbol{t}}^{(i)}$ (determined in the rough camera motion estimation stage) with respect to $\hat{\alpha}^{(i)}$

$$\Delta\tilde{\boldsymbol{t}}_*^{(i)} = \hat{\alpha}^{(i)} \frac{\Delta\tilde{\boldsymbol{t}}^{(i)}}{|\Delta\tilde{\boldsymbol{t}}^{(i)}|}. \tag{4}$$

Next, the estimated motion $\Delta\tilde{\mathbf{Q}}_*^{(i)}$ of the bronchoscope camera between frames $(i-1)$ and i can be computed by

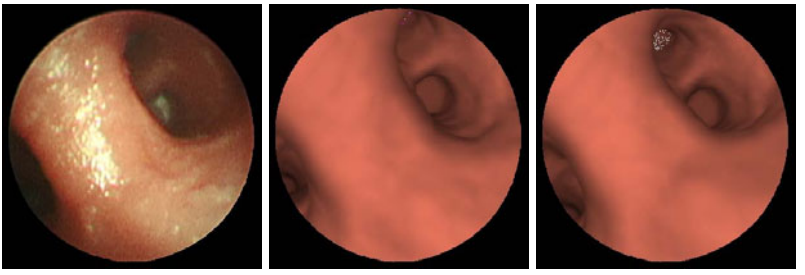
$$\Delta\tilde{\mathbf{Q}}_*^{(i)} = \begin{pmatrix} \Delta\tilde{\mathbf{R}}^{(i)} & \Delta\tilde{\boldsymbol{t}}_*^{(i)} \\ \mathbf{0}^T & 1 \end{pmatrix}, \tag{5}$$

where $\Delta\tilde{\mathbf{R}}^{(i)}$ is calculated by Eq. 2. Finally, the estimate $\Delta\tilde{\mathbf{Q}}_*^{(i)}$ is utilized as initialization of image registration, as described in the next step.

[Step 5] Intensity-based image registration. Intensity-based registration commonly defines a similarity measure and maximizes the similarities or minimizes the dissimilarities between an RB image $\mathbf{I}_R^{(i)}$ and a virtual bronchoscopic (VB) image $\mathbf{I}_V$. We here use a modified mean squared error (*MoMSE*) [5] similarity measure. Let $\mathbf{I}_V(\mathbf{Q}^{(i)})$ be a VB image generated from the predicted pose $\mathbf{Q}^{(i)} = \mathbf{Q}^{(i-1)}\Delta\mathbf{Q}^{(i)}$ of the current frame using volume rendering techniques, where $\mathbf{Q}^{(i-1)}$ denotes the previous camera pose and $\Delta\mathbf{Q}^{(i)}$ the inter-frame motion information between successive frames. By updating $\Delta\mathbf{Q}^{(i)}$, a series of VB images $\mathbf{I}_V(\mathbf{Q}^{(i-1)}\Delta\mathbf{Q}^{(i)})$ is generated and the most similar one corresponding to the RB image $\mathbf{I}_R^{(i)}$ is searched for. In summary, the intensity-based registration process optimizing $\Delta\mathbf{Q}^{(i)}$ can be formulated as

$$\Delta\mathbf{Q}^{(i)} = \arg\min_{\Delta\mathbf{Q}} MoMSE(\mathbf{I}_R^{(i)}, \mathbf{I}_V(\mathbf{Q}^{(i-1)}\Delta\mathbf{Q})). \tag{6}$$

For this optimization, the initialization of $\Delta\mathbf{Q}$ in Eq. 6 is one of the key components affecting tracking robustness and accuracy. $\Delta\mathbf{Q}$ is initialized as an identity matrix in previous work [5]. However, in our new method, we use our estimate $\Delta\tilde{\mathbf{Q}}_*^{(i)}$ (see Eq. 5) instead. Since we got this estimate by matching stable image features, it can overcome certain limitations of sole image registration such as dependencies on airway folds or bifurcations and hence enhances the tracking performance.



(a) **(b)** **(c)**

Fig. 2. Example of the tracking results from the two stages. (a) shows the real pose of the bronchoscope camera. (b) displays the predicted pose from rough camera motion estimation by using feature-based tracking. (c) shows the refined pose by performing image registration initialized by (b).

3 Experimental Results and Discussion

We evaluated sole intensity-based registration (M1) and our proposed tracking methods (M2: KLT-based method, M3: SURF-based method, M4: SIFT-based method) on patient datasets, each consisting of an RB video sequence and a preinterventional 3-D chest CT. In-vivo patient data was acquired in accordance with a standard clinical protocol. The acquisition parameters of the CT images are 512×512 pixels, 72-209 slices, 2.0-5.0 mm slice thickness, and 1.0-2.0 mm reconstruction pitch. The image sizes of the bronchoscopic video frames are 362×370 and 256×263 pixels. We have done all implementations on a Microsoft

Table 1. Comparison of the tracking results for our patient studies, in terms of the number and percentage of successfully tracked frames and average processing time (seconds) per frame

Cases ID	Num. of Frames	Number (Percentage) of frames successfully tracked			
		M1	M2 (KLT)	M3 (SIFT)	M4 (SURF)
Case 1	1200	450 (37.5%)	560 (46.7%)	683 (56.9%)	1120 (93.3%)
Case 2	200	116 (58.0%)	120 (60.0%)	70 (35.0%)	130 (65.0%)
Case 3	800	433 (54.1%)	618 (77.2%)	694 (86.8%)	774 (96.7%)
Case 4	800	437 (54.6%)	340 (42.5%)	605 (75.6%)	780 (97.5%)
Case 5	1000	431 (43.1%)	506 (50.6%)	575 (57.5%)	557 (55.7%)
Case 6	279	279 (100%)	279 (100%)	279 (100%)	279 (100%)
Case 7	400	240 (60.0%)	190 (32.5%)	210 (52.5%)	260 (65.0%)
Case 8	450	246 (54.7%)	217 (48.2%)	10 (2.22%)	10 (2.22%)
Total	**5120**	**2632 (51.4%)**	**2830 (55.3%)**	**3126 (61.1%)**	**3910 (76.4%)**
Average Times		**0.92 s**	**0.96 s**	**0.57 s**	**1.83 s**

Visual C++ platform and ran it on a conventional PC (CPU: Intel XEON 3.80 GHz×2 processors, 4-GByte memory).

A criterion for determining whether a method is more robust than another can be described by visual inspection and sum of the number of successfully tracked frames. If a VB image generated from the estimated camera parameters is greatly similar to the corresponding RB image, we consider it successfully tracked.

Table 1 gives quantitative results on the performance of all methods. Compared to M1, M2, and M3, in most cases the tracking performance has been improved significantly by using the proposed tracking algorithm M4. Figure 4 shows examples of RB images and the corresponding virtual images generated by volume rendering using the camera pose, calculated by the respective methods. The virtual images generated from the estimates of M4 are more similar than those of M1, M2, and M3, which means M4 more accurately predicts the real pose.

For the KLT method, we detect corner features and select 430 good features to be tracked from the previous frame [11]. The KLT tracker can usually track

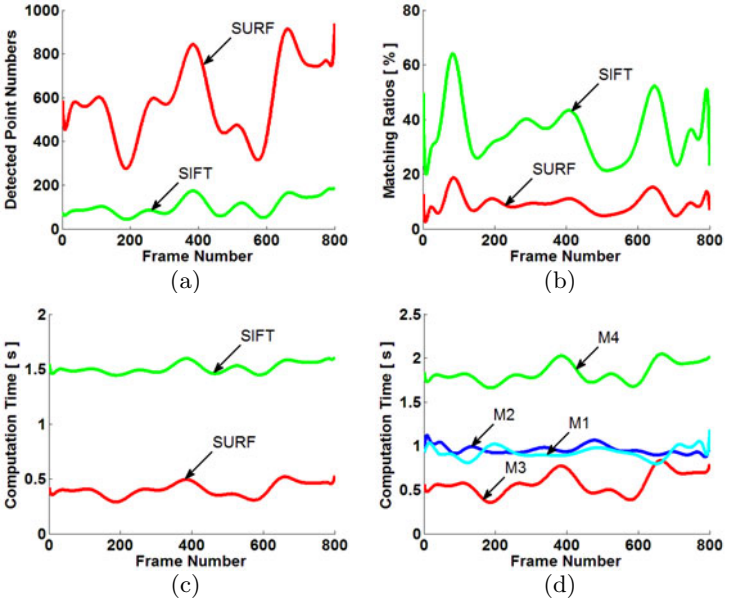


Fig. 3. Examples of detected feature numbers and computation times of Case 4. (a) shows the number of detected feature points, (b) gives the matching ratios calculated between the numbers of matching and detected points when using SURF and SIFT for each frame. (c) displays the time required for detecting SURF and SIFT features, and (d) illustrates the time needed to track the bronchoscope pose of each frame when using M1, M2, M3, and M4. It clearly shows that for M4 the average processing time with at least 1.5 seconds per frame is three times higher than that for M3, because it includes SIFT feature detection.

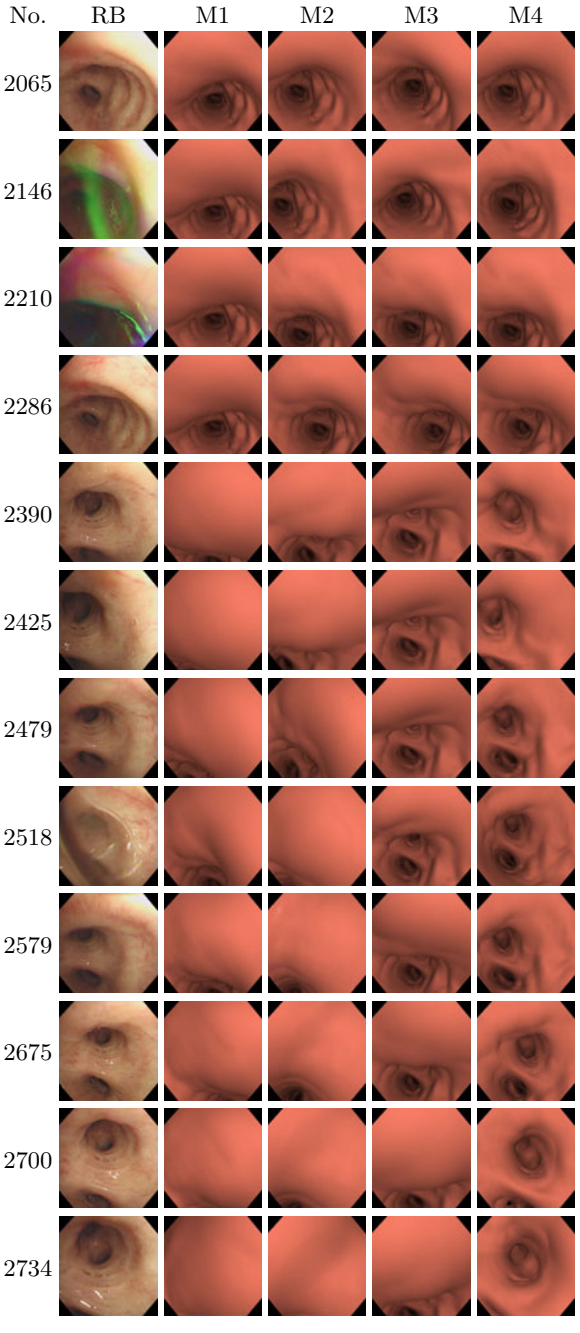


Fig. 4. Results of camera motion tracking for the patient assessment. The second column shows selected frames from a sequence of patient RB images and the first column their corresponding frame numbers. The other columns show tracking results for methods M1∼M4, all generated by volume rendering of the airways from the estimated viewpoints.

around 200 points per frame in our case. Because of the quality of KLT features, M2 has worse tracking results than M3 and M4, but is still better than M1. The tracking results of M3 are worse than those of M4, although SURF detected many more features and correspondences (around 587 detected points and 56 matching points per frame, matching ratios: 9.2%, as shown in Figure 3 (a) and (b)) than that of SIFT (around 103 detected points and 38 matching points per frame, matching ratios: 35.7%, as shown in Figure 3(a) and (b)). We believe that the SURF features-based method gives worse estimates to initialize the registration step than the SIFT features-based method, and hence fails to track the bronchoscope motion more often. This also demonstrates that the feature point quality from SURF is not as good as that of SIFT, as the authors already concluded in the work of Bay et al. [2]. Additionally, we note that all other approaches (M2-M4) show better tracking results than sole intensity-based image registration (M1). This can be explained by the usage of image texture features that depend less on airway folds or bifurcations.

Regarding computational efficiency, according to Table 1, M1 requires approximately 0.92 seconds to process a frame and the run-time of M2 is about 0.96 seconds per frame while that of M3 comes to 0.57 seconds per frame and M4 computes each frame in around 1.83 seconds. Compared to M1, M3 can improve the computational efficiency while M4 increases the processing time for each frame. From the work of Bay et al. [2] we know that SURF is faster than SIFT at detecting features, since the SURF method uses a fast-Hessian detector on the basis of an integral image. However, all methods cannot track the bronchoscope motion in real time (real time means 30 frames per second need to be processed in our case). This is because feature-based motion recovery methods are time-consuming in terms of detecting points and finding their correspondences, and so is the registration stage of bronchoscope tracking. However, we can utilize the GPU (graphics processing unit) to accelerate our implementations and make it (almost) real time.

Finally, in our patient study all methods failed to track the motion of the bronchoscope in some cases. This is because the estimation failed in the intensity-based registration process, which is usually caused by problematic bronchoscopic video frames such as RB images, on which bubbles appeared. Additionally, tracking failure also resulted from airways deformation, which was caused by patient movement, breathing, and coughing, and is also one particular challenge in navigated bronchoscopy. Currently, we do not explicitly address the problem of respiratory motion in our tracking method. Therefore, our future work will focus on improving intensity-based image registration for bronchoscope tracking during bronchoscopic navigation, as well as constructing a breathing motion model to compensate for respiratory motion.

4 Conclusion

This paper compared KLT, SURF, and SIFT features applied to bronchoscope tracking. We utilized the KLT-, SURF-, and SIFT-feature-based camera motion tracking method to improve the performance of image registration-based

bronchoscope tracking without an additional position sensor. Furthermore, from experimental results, we may conclude that SIFT features are more robust than the other two features when applied to predict bronchoscope motion, since the SIFT-based method successfully tracked 76.4% frames, compared to the KLT-based and the SURF-based methods with 55.3% and 61.1%, respectively. However, with about half to a third the processing time of the other methods, the SURF-based method seems to be a good compromise between tracking accuracy and computational efficiency.

References

1. Bauer, J., Süunderhauf, N., Protzel, P.: Comparing several implementations of two recently published feature detectors. In: Proc. of the International Conference on Intelligent and Autonomous Systems (2007)
2. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: Speeded-up robust features (surf). Computer Vision and Image Understanding 110(3), 346–359 (2008)
3. Cuevas, E., Zaldivar, D., Rojas, R.: Kalman filter for vision tracking. Tech. Rep. B 05-12, Freie University Berlin (October 2005)
4. Burschka, D., Li, M., Ishiia, M., Taylora, R.H., Hager, G.D.: Scale-invariant registration of monocular endoscopic images to CT-scans for sinus surgery. Medical Image Analysis 9(5), 413–426 (2005)
5. Deguchi, D., Mori, K., Feuerstein, M., Kitasaka, T., Maurer Jr., C.R., Suenaga, Y., Takabatake, H., Mori, M., Natori, H.: Selective image similarity measure for bronchoscope tracking based on image registration. Medical Image Analysis 13(4), 621–633 (2009)
6. Deligianni, F., Chung, A.J., Yang, G.Z.: Nonrigid 2-D/3-D registration for patient specific bronchoscopy simulation with statistical shape modeling: Phantom validation. IEEE Transactions on Medical Imaging 25(11), 1462–1471 (2006)
7. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge University Press, Cambridge (2004)
8. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
9. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(10), 1615–1630 (2005)
10. Mori, K., Deguchi, D., Sugiyama, J., Suenaga, Y., Toriwaki, J., Maurer Jr., 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. Medical Image Analysis 6, 321–336 (2002)
11. Shi, J., Tomasi, C.: Good features to track. In: CVPR, pp. 593–600 (1994)
12. Solomon, S.B., White Jr., P., Wiener, C.M., Orens, J.B., Wang, K.P.: Threedimensionsal CT-guided bronchoscopy with a real-time electromagnetic position sensor: a comparison of two image registration methods. Chest 118(6), 1783–1787 (2000)

Modeling Kinematics of Mobile C-Arm and Operating Table as an Integrated Six Degrees of Freedom Imaging System

Lejing Wang[1], Rui Zou[1], Simon Weidert[3], Juergen Landes[3], Ekkehard Euler[3], Darius Burschka[2], and Nassir Navab[1]

[1] Chair for Computer Aided Medical Procedures (CAMP), TU Munich, Germany
[2] Chair for Robotics and Embedded Systems, TU Munich, Germany
[3] Trauma Surgery Department, Klinikum Innenstadt, LMU Munich, Germany

Abstract. Maneuvering mobile C-arms to a desired position and orientation during surgery is not only a routine surgical task, e.g. for C-arm repositioning, but also an indispensable step for advanced X-ray imaging techniques, e.g. parallax-free X-ray image stitching. Standard mobile C-arms have only five degrees of freedom (DOF), which definitely restricts their motions that have six DOF in 3D Cartesian space. In this paper, we enable the mobile C-arm to have six DOF relative to the patient's table by integrating a translational movement of the patient's table into the mobile C-arm kinematics. We present a novel method to model the kinematics of the mobile C-arm and operating table as an integrated 6DOF C-arm X-ray imaging system. Kinematic singularities of the 6DOF C-arm model are determined by analyzing its manipulator Jacobian matrix. Inverse kinematic analysis is employed in order to find the required joint values to move the C-arm into the desired position and orientation. Our proposed 6DOF C-arm modeling paves the way for advanced applications in the fields of surgical navigation and advanced X-ray imaging that require C-arms to be precisely positioned or repositioned relative to the patient's table with six DOF. In our implementation, we employ a visual planar marker pattern and a standard mobile C-arm augmented by a video camera in order to obtain a relationship between the C-arm system and the patient's table. C-arm repositioning experiments on phantom study demonstrate the practicality and accuracy of our developed 6DOF C-arm system, and show the improved accuracy of C-arm repositioning by using the 6DOF C-arm model over the 5DOF C-arm model.

1 Introduction and Related Work

Modern trauma and orthopedic surgical procedures use X-ray images taken by mobile C-arms during surgery as intervention guidance, especially in minimally invasive surgery. Moving the mobile C-arm into the best viewing position in regard to the anatomy is a common surgical task, which requires experiences, time, and sometimes many X-ray shots until the desired image is obtained. Moreover, it is also an indispensable step for advanced X-ray imaging techniques,

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 485–493, 2010.

e.g. parallax-free X-ray image stitching [1] and intra-operative cone-beam CT [2]. However, positioning the mobile C-arm is mostly difficult and impractical to accomplish, since the X-ray source of the standard mobile C-arm is controlled by a complex kinematic chain of the five joints. Several methods were proposed to position a C-arm system without radiation, e.g. the visual serving based method in [3], the robotized mobile C-arm [2] and the inverse C-arm positioning using real-time body part detection [4].

Standard mobile C-arms only have five joints (five DOF), three of which are rotation joints and two are translation joints. As in 3D Cartesian space six DOF are required, the C-arm end-effector, i.e. the X-ray source, is certainly restricted in terms of reaching an arbitrary position and orientation. Matthaeus et al. presented the inverse kinematics of a general mobile C-arm, proving the existence of necessary joint parameters for imaging a given point from a given direction [2]. They reduced the 3D Cartesian space to five DOF by only considering a 2DOF direction instead of a 3DOF orientation. In their work about parallax-free X-ray image stitching [1], Wang et al. proposed a method of enabling the X-ray source to have a relative pure rotation to the patient's table, by moving the table as an additional DOF in order to compensate for the translational motion of the X-ray source. However, they did not include this DOF into the formulation.

In this work, we propose to integrate a translational movement of the patient's table into the mobile C-arm's kinematics, thus enabling the X-ray source to have six DOF with respect to the patient's table. We present a method to model the kinematics of the mobile C-arm and operating table as an integrated 6DOF C-arm X-ray imaging system. A forward kinematic analysis is performed in order to build a kinematic chain model for the 6DOF C-arm model. Kinematic singular configurations of the 6DOF C-arm model are determined by analyzing the manipulator Jacobian matrix. The inverse kinematics is solved to position the X-ray source at an arbitrary position and orientation relative to the patient's table. In our implementation, we employ a visual planar marker pattern and a Camera Augmented Mobile C-arm (CamC) system, which is a standard mobile C-arm augmented by a video camera and mirror construction, in order to obtain a relationship between the C-arm system and the patient's table . However, this can also be accomplished by using the X-ray marker based C-arm pose estimation method of [5] or external tracking systems. Being able to position the C-arm relative to the patient with full 6DOF using C-arms which have 5DOF paves the way for many computer assisted clinical applications. For example, precisely repositioning the C-arm during surgery as shown in our experiment. In particular, this 6DOF C-arm model can allow automatic C-arm positioning for parallax-free X-ray image stitching proposed in [1]. In their work, the authors move the table as an additional DOF in order to enable the X-ray source to have a relative pure rotation for generating parallax-free panoramic X-ray images. However, they did not include this DOF into the formulation. Moreover, having a 6DOF C-arm kinematic model would in general enable the smooth transform of the CT-based pre-operative planning into the operating room, by accurately positioning the patient with regard to the X-ray C-arm.

2 Method

The standard mobile C-arm has five joint parameters: the vertical, which translates the C-arm up and down along the axis Z_1; the wigwag (or swivel), which rotates the "C" around the axis Z_2; the horizontal, which changes the arm length along the axis Z_3; the angular, which rotates the "C" around its center axis Z_4; the orbital, which rotates the "C" in its own plane around the axis Z_5 (see figure 2). Defining a unique pose in 3D Cartesian space requires six independent parameters, i.e. three rotation parameters for the orientation and three translation parameters for the position. It is impossible for mobile C-arms that have only five DOF to satisfy an arbitrary X-ray source pose. Therefore, we propose to integrate a translational movement of the patient's table into the C-arm kinematics in order to enable the X-ray source to be positioned at an arbitrary pose relative to the operating table. Note that, this does not take into account the limited mechanical range of each joint.

2.1 System Components

Our system is composed of a translatable operating table, a visual planar square marker pattern and a CamC system built by attaching a video camera and mirror construction to a mobile C-arm, Siremobile Iso-C 3D C-arm (see figure 1). The CamC system is calibrated by using the proposed method of [6] to enable that the X-ray source and the video camera have the same projection geometry. We designed a visual planar marker pattern, in which all the square markers can be uniquely distinguished [7]. The marker pattern is printed in A2 size paper by a high definition printer, and is rigidly and flatly attached under the operating table.

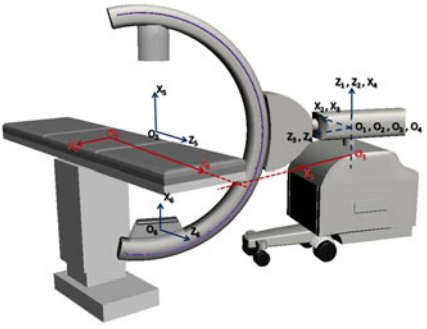
Fig. 1. System components for the implementation of the 6DOF C-arm system. The right top image shows the attachment of the visual marker pattern under the table. The right bottom image shows our custom made camera and mirror construction.

2.2 Integrated 6DOF C-Arm System Kinematic Modeling

The forward kinematic analysis of the 5DOF C-arm is performed to build a direct relation from the five C-arm joint values to the pose of the X-ray source. We assign a coordinate frame to each link according to Denavit-Hartenberg (DH) rules [8]. The origin of the coordinate frame 1 is chosen at the C-arm base. The origin of the coordinate frame 6 is chosen at the X-ray source center, which is the end-effector. Figure 2 shows the assigned coordinate frames marked by black color for the 5DOF C-arm kinematic model. Note that this kinematic model is based on isocentric C-arms, but with a minor modification it can also be applied to non-isocentric C-arms. Similar forward kinematic analysis of the C-arm has been studied in [2,3]. From the assignment of the coordinate frames, the transformation matrix ${}^{1}T_{6} \in \mathbb{R}^{4\times4}$ from the X-ray source to the C-arm base coordinate frame can be established according to the method of [9]. ${}^{1}T_{6}$ is the forward kinematic function of the five joint variables for the 5DOF C-arm model.

We integrate the forward-backward translation of the table into the 5DOF C-arm kinematics in order to build a 6DOF C-arm X-ray imaging model. Translating the table forward-backward is equivalent to moving the whole C-arm system in an opposite direction. Thus, the table translation must be the first joint in the 6DOF C-arm kinematic model. The coordinate frame 0 assigned to the table is defined as, the Z_0-axis parallel to the direction of the table forward-backward translation, the X_0-axis chosen such that the X_0-Z_0 plane is parallel to the table surface, and the Y_0-axis following the right hand rule (see figure 2). C-arm pose estimation is performed to get a transformation ${}^{6}H_{0} \in \mathbb{R}^{4\times4}$ from the patient's table to the C-arm X-ray source. We employ the attached video camera of the CamC system and the visual marker pattern for C-arm pose estimation without radiation. The corners of each square marker in the marker pattern can be extracted with subpixel accuracy and used as feature points. Having the marker pattern with known geometry, we are able to establish point correspondences between the 2D image points and 3D space points. Based on these point correspondences, the pose of the camera (X-ray source) relative to the marker pattern (patient's table) is computed by using a standard camera calibration method of [10]. Thus, we can compute a transformation from the table to the C-arm base, ${}^{1}H_{0} = {}^{1}T_{6} \cdot {}^{6}H_{0}$.

We develop a forward kinematic chain for the 6DOF C-arm model by defining the table coordinate frame as the base coordinate frame and coupling it to the 5DOF C-arm kinematic chain model. In order to satisfy the DH rules, we only need to re-assign the coordinate frame 1 of the 5DOF C-arm model. The vertical movement axis Z_1 remains the same as in the 5DOF model. According to the DH rules, the axis X_1 should be re-assigned such that it is orthogonal to and intersects with both axes Z_0 and Z_1. With the known transformation between the axes Z_0 and Z_1, we look for two points on both axes, which define a minimum distance between these two axes. The axis X_1 is defined by these two points. Figure 2 shows the coordinate frames of the 6DOF C-arm kinematic model, in which coordinate frames 0 and 1 are marked by red, and coordinate frames 2-6 are marked by black. After establishing all the six coordinate frames, required



(a) Coordinate frames assignment

	a (Tx)	d (Tz)	α (Rx)	θ(Rz)
0	Dist(Z_0,Z_1)	d_0^* + Dist(X_0,X_1)	Ang(Z_0,Z_1)	Ang(X_0,X_1)
1	0	d_1^* + Dist(X_1,X_2)	0	Ang(X_1,X_2)
2	0	0	-pi/2	θ_2^*
3	0	d_3^*	0	-pi/2
4	0	length_offset	pi/2	θ_4^*
5	-orbital_offset	0	0	θ_5^*

(b) Link parameters table(*variable)

Fig. 2. (a) Coordinate frames assigned for the 5DOF C-arm and the integrated 6DOF C-arm model according to the Denavit-Hartenberg rules; (b) the link parameters table of the 6DOF C-arm model. $Ang(A, B)$(or $Dist(A, B)$) represents the angle(or distance) between axes A and B.

link parameters for building the forward kinematic function are obtained and shown in Figure 2. Let $q_{6dof} = [d_0; d_1; \theta_2; d_3; \theta_4; \theta_5]^T$ be a vector containing the six joint variables. A transformation matrix ${}^iA_{i+1} \in \mathbb{R}^{4\times4}$ from the coordinate frame i+1 to i can be derived from the link parameters [9]. A forward kinematic function ${}^0T_6(q_{6dof})$ for the 6DOF C-arm model is established by: ${}^0T_6(q_{6dof}) = {}^0A_1 \cdot {}^1A_2 \cdot {}^2A_3 \cdot {}^3A_4 \cdot {}^4A_5 \cdot {}^5A_6$, which represents the pose of the X-ray source with respect to the patient's table.

2.3 Kinematic Singularity

Kinematic singularity is a kinematic configuration, at which the mobility of the end-effector is reduced, i.e. losing one or more DOF of motion. A kinematic system has a kinematic singularity for a specific joint configuration when the rank of its manipulator Jacobian matrix is less than the number of the required DOF. We derive the manipulator Jacobian matrix J_{6dof} for the 6DOF C-arm model from its forward kinematic chain model according to the method proposed in [9]. J_{6dof} is a function of variable q_{6dof} and relates differential changes in the six joint positions to the X-ray source linear and angular velocity. The 6DOF C-arm model consists of six joints, and thus J_{6dof} is a 6x6 square matrix. Therefore, when the determinant of J_{6dof} is zero, the rank of J_{6dof} becomes less than six. We compute the determinant of J_{6dof},

$$det(J_{6dof}) = (sin(Ang(X_1, X_2))cos(\theta_2) + cos(Ang(X_1, X_2))sin(\theta_2)) \cdot sin(Ang(Z_0, Z_1)) \cdot cos(\theta_4) \quad (1)$$

$det(J_{6dof}) = 0$, when

a) $sin(Ang(Z_0, Z_1)) = 0$, the direction of the table translation is parallel to the C-arm vertical movement direction. This situation does not physically exist.
b) $cos(\theta_4) = 0$, the orbital rotation axis Z_5 is parallel to the wigwag rotation axis Z_2. Consequently, one DOF is missing.
c) $sin(Ang(X_1, X_2))cos(\theta_2) + cos(Ang(X_1, X_2))sin(\theta_2) = 0 \Rightarrow \theta_2 = -Ang(X_1, X_2)$, the direction of the table translation is parallel to the C-arm horizontal movement direction. Therefore, this configuration leads to losing one DOF.

(a) (b) (c)

Fig. 3. Example positions for the kinematic singularities of the 6DOF C-arm model. (a) $sin(Ang(Z_0, Z_1)) = 0$; (b) $cos(\theta_4) = 0$; (c) $\theta_2 = -Ang(X_1, X_2)$.

2.4 Controlling of C-Arm X-Ray Source

Controlling of the C-arm requires finding the required joint values q to position the C-arm X-ray source to a desired pose T, which is an inverse kinematic problem. Currently, we employ Newton iterative method of using pseudo-inverse Jacobian [11] to solve the inverse kinematic problem for both our developed 6DOF C-arm model and standard 5DOF C-arm model. Let J^+ be the pseudo-inverse Jacobian, F be the forward kinematic function, and $\Delta(F(q_n) - T)$ represent the difference between two rigid transformations as a six-vector of displacements and rotations. The calculation is repeated as equation 2 until $\parallel J^+(q_n)\Delta(F(q_n) - T) \parallel$ reaches a pre-defined minimun value or a maximum iteration number is reached.

$$q_{n+1} = q_n + J^+(q_n)\Delta(F(q_n) - T) \tag{2}$$

Using pseudo-inverse Jacobian allows us to obtain a solution for the 5DOF C-arm model and for kinematic singularities of the 6DOF C-arm model. However, we cannot get a converged solution when unfeasible movement directions are involved.

3 Experiment

Intra-operative repositioning of mobile C-arms is a common surgical task. During the intervention, the C-arm often has to be moved back to acquire the second X-ray image from the same viewing point as the first one, e.g. for the confirmation of surgical outcome. We evaluated the practicality and accuracy of our developed

6DOF C-arm system by performing C-arm repositioning tasks. Moreover, the experiments also show the improved accuracy of C-arm repositioning by using the 6DOF C-arm model over the 5DOF C-arm model.

For one repositioning experiment, the C-arm was first positioned to a reference position, at which the first X-ray image was acquired. Then, the operating table was translated and the C-arm was moved to a starting position, which is a C-arm zero-joint configuration, i.e. all the five C-arm joints are set to zero. The moved joints are shown in the column "actual moved joints" in table 1. At the starting position, the 6DOF C-arm kinematic model was built by using our proposed modeling method. After this, the required joint movements for repositioning were computed. Thanks to the visual marker based C-arm pose estimation, we are able to compute the current joint values for each estimated C-arm pose by solving the inverse kinematics. Therefore, our system provides a continuous guidance for the operator of the device how to move each joint (including the table translation) in order to move the C-arm to the pre-defined reference position. Then, the second X-ray image was acquired after the C-arm was repositioned. The C-arm was considered to be repositioned if further required movements are below 1mm for translational joints and 0.5^o for rotational joints. We evaluate the quality of C-arm repositioning by analyzing the image difference between the first and the second X-ray images. Four X-ray visible square markers that can be uniquely detected in X-ray images are placed on the operating table. The corners of each marker are extracted in X-ray images with subpixel accuracy and used to compute the image difference, which is defined as the pixel-distance between corresponding corners.

We conducted five pairs of C-arm repositioning. Each pair, which had the same starting and reference positions, consisted of two repositioning procedures using our developed 6DOF C-arm model and the 5DOF C-arm model respectively. The pixel differences between the image acquired at the pre-defined position and the image acquired after repositioning are defined as errors. The error has two components: inverse kinematics error and system dependent errors (including the calibration of the CamC system, pose estimation and manual movement of C-arm). The inverse kinematics for 5DOF C-arm obtains an approximate solution and thus introduces errors, when an unfeasible movement of the X-ray source would be required. System dependent errors have a similar influence in the results of the repositioning tests using both 6DOF and 5DOF C-arm models. Experimental results (see table 1) show that all the tests using 6DOF C-arm have relatively small and constant errors. The mean error of the five repositioning experiments using the 6DOF C-arm model is 5.42 ± 2.46 pixels. This demonstrates the accuracy of our developed 6DOF C-arm system. The error analysis for each test pair shows the improved accuracy of C-arm repositioning using the 6DOF C-arm over the 5DOF C-arm. Each test pair was performed once, since the joint motions for each test pair were randomized and could not be exactly reproduced. However, five test pairs were conducted to ensure the repeatability of the experiments.

Table 1. Results of C-arm repositioning experiments

Actual moved joints	Kinematic model	Computed joint movement values for repositioning: translation(mm), rotation(degree)						Image difference: Mean(pixels)
		table	vertical	wigwag	horizontal	angular	orbital	
table, angular	6DOF	102.15	-1.01	-0.13	0.81	5.52	0.32	7.35
	5DOF	non	-1.53	3.45	-1.38	3.89	0.03	70.16
table, wigwag	6DOF	62.07	-0.39	-6.48	-1.00	-0.16	-0.07	5.78
	5DOF	non	-0.45	-4.31	-5.10	-1.16	-0.5	64.40
table, orbital	6DOF	-73.22	-0.49	-0.001	-2.59	-0.06	-9.69	1.38
	5DOF	non	-0.64	-2.70	1.63	1.27	-9.60	59.99
table, horizontal, vertical, angular, orbital, wigwag	6DOF	31.04	43.15	-7.34	39.89	15.40	-16.05	5.18
	5DOF	non	41.70	-6.17	38.58	14.69	-15.41	56.43
vertical, angular, horizontal, orbital, wigwag	6DOF	8.61	42.13	-6.10	67.97	15.12	-7.25	7.41
	5DOF	non	41.84	-5.79	67.98	14.99	-7.33	8.21

4 Discussion and Conclusion

Standard mobile C-arms have only five DOF, which definitely restricts their motions that have six DOF in 3D Cartesian space. In this paper, we proposed to build a 6DOF C-arm X-ray imaging model by integrating a translational movement of the patient's table to the 5DOF C-arm kinematics. We presented a method to develop a forward kinematic chain for the 6DOF C-arm model without constraints on the initial setup of the table position, whose kinematic singularities were shown by analyzing the manipulator Jacobian. An iterative method of using pseudo-inverse Jacobian was applied to solve the inverse kinematics in order to position the X-ray source at an arbitrary pose relative to the patient's table. In our implementation, we employed the CamC system and the visual planar marker pattern in order to register the table translation into the mobile C-arm kinematics with no radiation. However, this can also be accomplished by using the X-ray marker based C-arm pose estimation method of [5] or external tracking systems. The C-arm repositioning task was chosen to evaluate the practicability and accuracy of our developed 6DOF C-arm system, and to show the advantages of using the 6DOF C-arm model over the 5DOF C-arm model. Our medical partners confirmed that, with the continuous instruction of how much to move each joint, the C-arm can be moved back to a pre-defined position within an acceptable accuracy and time. Currently, surgeons need to put additional efforts into manually moving the table and C-arm joints. However, motorized C-arms and operating tables are available in many clinical sites, which could automate the control of the 6DOF C-arm system. In their valuable work on inverse kinematics for a general 5DOF mobile C-arm, Matthaeus et al. [2] reduced 3D Cartesian space to five DOF by only considering a 2DOF direction instead of a 3DOF orientation. They rotate X-ray images around the principle point in order to compensate for the missing degree of freedom. However, in spite of its extreme usefulness in practice, the limited DOF could impose some functional constraints, e.g. obstacle avoidance, physical limited joint range, and singularity avoidance. Integrating a patient's table translation into the 5DOF C-arm to build a 6DOF C-arm model can definitely reduce the problems of the

functional constraints. Our proposed 6DOF C-arm modeling paves the way for advanced applications in the fields of surgical navigation and advanced X-ray imaging that require C-arms to be precisely positioned or repositioned relative to the patient's table with six DOF.

References

1. Wang, L., Traub, J., Weidert, S., Heining, S.M., Euler, E., Navab, N.: Parallax-free long bone x-ray image stitching. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 173–180. Springer, Heidelberg (2009)
2. Matthaeus, L., Binder, N., Bodensteiner, C., Schweikard, A.: Closed-form inverse kinematic solution for fluoroscopic c-arms. Advanced Robotics 21(8), 869–886 (2007)
3. Navab, N., Wiesner, S., Benhimane, S., Euler, E., Heining, S.M.: Visual servoing for intraoperative positioning and repositioning of mobile c-arms. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 551–560. Springer, Heidelberg (2006)
4. Schaller, C., Rohkohl, C., Penne, J., Stürmer, M., Hornegger, J.: Inverse c-arm positioning for interventional procedures using real-time body part detection. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 549–556. Springer, Heidelberg (2009)
5. Kainz, B., Grabner, M., Rüther, M.: Fast marker based c-arm pose estimation. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 652–659. Springer, Heidelberg (2008)
6. Navab, N., Mitschke, M., Bani-Hashemi, A.: Merging visible and invisible: Two camera-augmented mobile C-arm (CAMC) applications. In: Proc. IEEE and ACM Int'l Workshop on Augmented Reality, San Francisco, CA, USA, pp. 134–141 (1999)
7. Zhang, X., Fronz, S., Navab, N.: Visual marker detection and decoding in ar systems: A comparative study. In: IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2002) (October 2002)
8. Denavit, J., Hartenberg, R.S.: A kinematic notation for lower-pair mechanisms based on matrices. Trans. ASME J. Appl. Mech. 23, 215–221 (1955)
9. Spong, M.W., Hutchinson, S., Vidyasagar, M.: Robot Dynamics and Control, 2nd edn (2004)
10. Zhang, Z.: A flexible new technique for camera calibration. Technical report, Microsoft Research (1998) (last updated in 2008)
11. Chiaverini, S., Sciavicco, L., Siciliano, B.: Control of robotic systems through singularities. In: Advanced Robot Control, pp. 285–295. Springer, Heidelberg (1991)

Peripheral Lung Cancer Detection by Vascular Tumor Labeling Using *In-Vivo* Microendoscopy under Real Time 3D CT Image Guided Intervention

Miguel Valdivia y Alvarado, Tian Cheng He, Zhong Xue,
Stephen Wong, and Kelvin Wong*

The Center for Bioengineering and Informatics, Department of Radiology,
The Methodist Hospital Research Institute, Weill Cornell Medical College,
Houston, TX
kwong@tmhs.org

Abstract. We designed and evaluated a real time 3D CT Image Guided Intervention system that integrates in-vivo microendoscopic imaging for on-the-spot visualization of ICG contrast uptake by tumor vessel in peripheral lung tumors. The performance of the system was evaluated in seven rabbits where VX2 cells were implanted in the chest to create peripheral lung tumors. Two weeks later the animal underwent a chest CT scan which was used for creating a real time 3D vision and navigation tracking. ICG was injected fifteen minutes prior to the needle puncture to allow adequate contrast leakage inside the tumor and plasma clearance. After the needle puncture, the microendoscope was introduced inside the tumor for imaging. Visualization of tumor leaky vasculature was possible in all the tumors. The experiment demonstrated that real-time microendoscopy of deep solid organs under a 3D CT image-guided system is possible while providing enough accuracy in reaching tumors without complications.

Keywords: Microendoscopy, Real Time 3D CT Image Guided Intervention, Indocyanine Green (ICG), Tumor Leaky Vasculature, Lung Cancer Detection.

1 Introduction

Currently lung cancer has one of the highest mortality and morbidity rates, despite the efforts and investments made in new detection and treatment methods the long-term survival rate of lung cancer does not improve. The key to improve the long-term survival rate relies in early diagnosis, accurate localization and novel targeted therapies [1] for which new imaging techniques are expected to play a significant role**.** Among the news imaging techniques microendoscopy seems very promising as it allows direct observation of pathologic changes at the microscopic level, moreover fibered microscoendoscopy provides a clear, in-focus image of a thin section within a biological sample, [2] and it is capable of imaging fluorescent labeled biological

* Corresponding author.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 494–502, 2010.

structures *in vivo* at exceptionally high spatial resolutions. These technologies can be translated to the thoracic oncology field by assisting in early detection of precancerous and cancer conditions through improved biopsy selection. Considering that microendoscopic probes are able to image tissue with unprecedented spatial (less than 1 micron lateral resolution) and high temporal resolution (usually video rate) using low-cost, portable devices, this technology has the potential to decrease mortality and morbidity by optimizing lung tumor detection.

A major limitation of microendoscopy is the capacity to reach deep seated targets such as the lung [3]. To that end our group has developed an image guided intervention based on real time 3D CT images [4] and coupled with electromagnetic tracking system to enable microendoscopic probes to reach deep seated-targets. The integration of microscopic imaging with macroscopic image navigation will allow a closer look to small targets [3, 4], and may improve the accuracy of biopsies. Here we report the results of a study where we coupled microendoscopic imaging with the image guided intervention system we created for peripheral lung cancer detection. We evaluated its utility in small peripheral lung tumor diagnosis by validating the insertion accuracy of the 3D CT image guided intervention system in thoracic percutaneous punctures by using in vivo fibered microendoscopy for tumor detection using vascular labeling with Indocyanine Green contrast.

2 Materials and Methods

2.1 Microendoscopy Probes

A CellVizio Lung microendoscopic imaging system was used in the study. The CellVizio Lung system is a commercial system approved by the FDA for human lung application. This system offers a depth of observation between 0-50 µm, a lateral resolution of 3.5 µm, and excitation wavelength between 488 to 600 µm. The field of view ranges from 500 µm to 600 µm, covering a relatively large area of tissue and capturing optical image sequence of the fine tumor vasculature [5]. The video images were obtained at a rate of 12 frames per second. The manufacturer of the CellVizio system, Mauna Kea Technologies, provides us with a 1 mm diameter probe that allows using a smaller needle size for a percutaneous thoracic puncture, this probe will pass through the needle cover after the puncture is successfully achieved and the needle is inside the tumor.

2.2 Contrast Agents

Among the various contrast agents available Indocyanine Green (ICG) was selected because it is the only contrast agent approved by FDA for vasculature imaging; therefore a positive result can be translatable to the patient bedside faster. ICG is a sterile solution of a nontoxic tricarbocyanine dye with a peak spectral absorption of 790 nm, [6]. Experiments using near infra-red (NIR) reflected light have showed that it is possible to image sub-cellular features in the epithelial tissue with depth exceeding 400 µm. [7] NIR provides quantitative functional information that cannot be obtained by conventional radiological methods and also can provide *in vivo* measurements of the oxygenation and vascularization states. ICG is also a blood

pooling agent and has different delivery behavior between normal and cancer vasculature. In normal tissue, ICG acts as a blood flow indicator in tight capillaries of normal vessels. In tumors, ICG acts as a diffusible (extravascular) flow in leaky capillary of vessels.

2.3 3D CT Real-Time-Image Guided Intervention

Our image-guided intervention system consists of the following components (see Figure 1): An electromagnetic (EM) tracking device for real-time tracking of the needle introducer (Aurora Electromagnetic Measurement System); coherent software for real-time localization and visualization of multiple devices being tracked on the intra-procedural CT images and one fine needle (Traxtal Percunav, Biopsy Introducer 18G) for percutaneous lung puncture. The workflow of the system was implemented by using a modular multimodality image guidance platform (MIMIG) [8]. Diagnosis of the tumor mass inside the chest was accomplished based on morphologic and molecular imaging information obtained from different imaging modalities. Novel image computing tools including segmentation, registration, and microendoscopy image sequence processing were developed for MIMIG. The pre-procedural images were segmented for better visualization during surgical planning, and since fast segmentation is needed for intra-procedural images, the segmentation results were transformed onto the intra-procedural images for visualization during the intervention. Using these tools fast and accurate image segmentation and registration algorithms were developed for better visualization during surgical planning, intervention, and alignment of pre-procedural and intra-procedural images.

2.4 Animal Model

The animal protocol was revised and approved by the Comparative Medicine Program and IACUC committee at our institution. Solitary lung tumors in seven White New

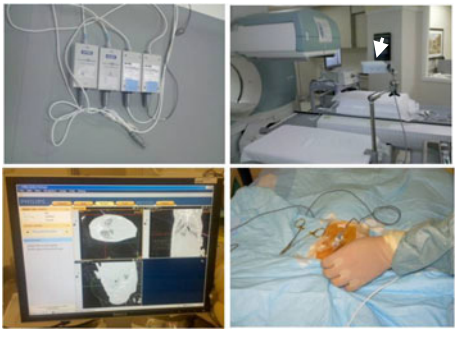


Fig. 1. Pictures showing various components of the real time 3D CT Image Guided Navigation System developed by our group: the upper left picture displays the connection trackers for EM tracking, the upper right picture shows the field generator that tracks the needle (white arrow). The bottom left picture shows the 3D image display available for the physician to design a pre-operative path planning in order to guide the needle to the tumor, and the bottom right picture shows the needle insertion inside the animal chest.

Zealand rabbits (2.0-2.2 kg) using VX2 carcinoma cells were created for the study. The first cell line was obtained from our cancer research center partner and injected directly on the limbs of one rabbit. The tumors first grew in the hind legs of a carrier rabbit for two weeks. This animal was sacrificed and the VX2 tumors were harvested from the legs to create a cell suspension at a density of $8x10^6$ cells/mL.

From two VX2 tumors three needles with 0, 5 ml cell suspension were created, two needles were injected into the limbs of another rabbit for tumor line preservation and the third one was administered in one animal tumor recipient. The cell suspension was administered in the right lower lung of the recipient rabbits using a 22-gauge Chiba needle percutaneously inserted using fluoroscopy. [9, 10] After the fifth day from the lung inoculation a CT scanning was performed weekly to follow the growth of the VX2 tumor. Tumors of ≤15 mm in diameter were confirmed at days 12-14 after inoculation in all the animals' models.

2.5 Experiments

Once the animal model has the desired tumor size it was anesthetized and transported to the CT room for the image-guided procedure. First, a pre-procedural CT scan of the thorax was obtained using a SPECT/CT (Symbia TruePoint from Siemens) with the following parameters: 1.25mm helical acquisition, pitch of 1.55 and a reconstruction of 1.2mm. The chest scan information was transferred from the CT scanner to the Image Guided Intervention system where the physician chooses a path planning for the percutaneous needle puncture according to the tumor location. The real time 3D vision and tool tracking was possible superimposing the electromagnetic tracking data (information about the instruments location) to the pre-procedural scan in real time, thus the display appears modulated by the real-time tracking data. The percutaneous accuracy puncture accuracy was later validated by coregistering pre- and post-punctures CT images.

Indocyanine Green (ICG-125 mg) contrast was injected IV 15 minutes before the percutaneous thoracic puncture; this period was set as the time necessary for adequate contrast leakage inside the tumor and contrast systemic clearance from the plasma. Before the puncture the mechanical ventilation was stopped to decrease the respiratory movement to the minimum. Once the tumor target was reached the needle was fixated and a post-procedural scan was performed to verify the needle location inside the tumor. After the image-guided needle puncture, the needle was retracted, leaving the needle introducer in place. Then, the 1mm O.D. Cellvizio microendoscope was inserted through the needle introducer and video recording from the VX2 tumors vessels started (Fig.2). At the end of the video recording a tumor biopsy was taken using a Biopsy Needle (Quick-Core, Cook Medical) that pass through the same needle introducer used for tumor imaging. After completing the biopsy the animals were euthanized, the chest was opened and the tumors were recovered. Samples from the needle biopsy and tumor were sent to histological analysis by a research vet tech blinded to the tumor source.

For detecting the ICG expression in the VX2 tumor vessels a pre-defined threshold of two-fold specific labeling over background were used. With these thresholds, the in situ ICG vessel expression inside the tumor can be visualized in real time. The percentage of tumor vessel expressing high uptake of ICG per frame was calculated

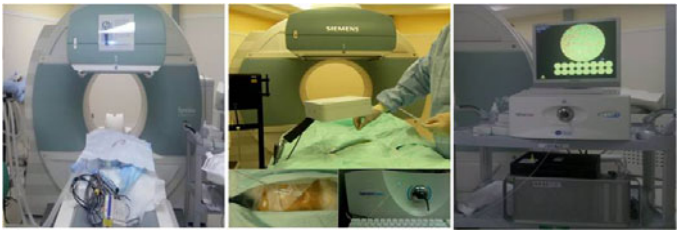


Fig. 2. In the left picture the animal model is undergoing the pre-procedural scan, the picture in the middle shows the introduction of the Cellvizio microendoscope probe inside the VX2 tumor and finally the right picture shows the imaging of tumor vessels labeled with Indocyanine Green.

after doing a post-processing with the pre-defined threshold. To increase reliability motion correction and scene change detection were applied along the video sequences. The ICG fluorescence microendoscopy was validated by semi-quantitative data analysis of fluorescence video intensity in normal lung and tumor tissues at different regions of the lung. The tumor was sampled 5 times while the normal tissue was sampled 2 times due to the low standard deviation of mean fluorescence video intensity in normal lung tissue and larger variation in tumor tissues.

3 Results

The tumors grew to a size equal or less than 15 mm measured by CT scan (average 13 mm) in the seven animal models used in the study. Complications related to thoracic puncture such as pneumothorax were not seen in the second scan. In the confirmatory scan the needle tip reached the VX2 tumors in the seven experiments, therefore it was not necessary to adjust the needle once inside the thorax. The needle tip was seen in the center of the tumor in the post-procedural CT scan in five animals. In the remaining two animals the needle tip was located in a peripheral zone of the tumor. Four to six videos were recorded in each animal and each video had a duration period between 35 to 55 seconds.

A higher contrast was expected because the micron-sized resolution and shadow reception of the microendoscope removed the depth dependence and partial volume effect. Figure 3 shows the fluorescent microendoscopic images recorded during the experiments, according to an intensity color map. In the left side of figure 3 we can appreciate frames from a video recorded with the needle tip in the peripheral zone of the VX2 tumor while in the right side frames from a centrally located needle tip case are showed. The first frame shows the original fluorescent image, and the next one corresponds to the segmented fluorescent image. The tumors vessels have a heterogeneous and high expression/uptake of ICG compared to areas with less vasculature, which had an intensity value between 40-60. This was particularly evident in the experiments where the needle tip was in a peripheral location, which is a zone less vascular compared to the videos recorded from the tumor center where the uptake of ICG was higher, due to the high angiogenesis rate. Also some dark areas were seen in the videos corresponding to necrotic areas due to hypoxia.

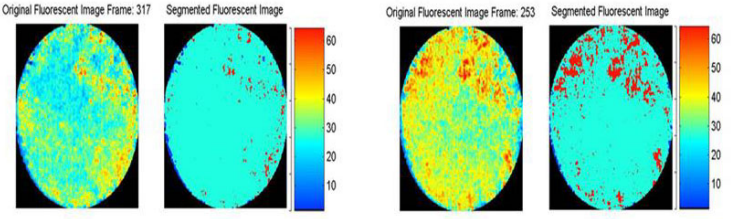


Fig. 3. Images captured from a VX2 rabbit model experiment cases where the needle tip had a peripheral location (left) and a case where the needle tip had a central location (right) inside the VX2 tumor. Note the stronger vascular labeling and expression obtained from the tumor with a central needle tip location (zones labeled with red and yellow colors).

A more dramatic example about the angiogenesis phenomena can be appreciated in figure 4 where the video sequences recorded showed regions with high uptake of contrast at the center of the tumor; the contrast intensity decreased gradually when the needle and microendoscope were pulled-back to a peripheral zone of the tumor and later disappeared when the microendoscope was outside the tumor (needle position confirmed with CT scan).

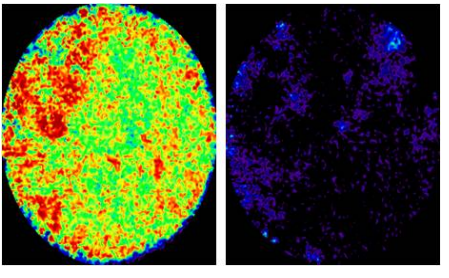


Fig. 4. In vivo microendoscopy images showing ICG-labeled vascular architecture inside the VX2 tumor first and later structure of the normal parenchyma when the needle was pulled back from the tumor

The biopsies taken at the end of the procedure had a positive result for VX2 carcinoma, which was compared to samples of the tumor excised from the lung at the end of the procedure (Figure 5).

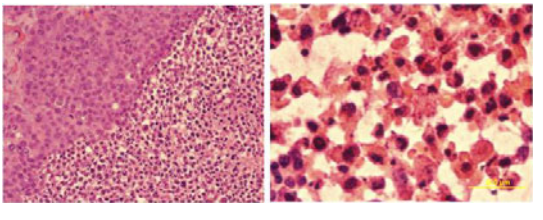


Fig. 5. The first slide was made from a sample of the recovered tumor after euthanasia, while the second slide comes from the sample obtained through the biopsy (both are from the same tumor). In both samples cells corresponding to VX2 carcinoma were reported.

4 Discussion

In chest medicine, microendoscopy technology has been applied in bronchoscopic procedures previously [11], where it was possible to obtain high-quality images through a fiber-optic probe of 1-mm diameter introduced in the working channel of a bronchoscope. Even more, image registration for displaying the position of flexible bronchoscope on CT scan during transbronchial bronchoscopy needle aspiration has been applied successfully [12, 13]. However, despite its effectiveness, the bronchoscopic approach is useless as a diagnosis or treatment tool for tumors outside the airways, therefore not an option for peripheral lung tumors. For this kind of tumors the work-up plan usually includes a diagnosis made through a percutaneous biopsy and the treatment is either performed with surgery or ablation. Using the real time 3D CT Image Intervention system microendoscopic imaging for tumor diagnosis can be performed through a percutanoues biopsy needle as we proved in our study.

The time necessary to do the thoracic puncture with the help of this system, ICG injection and clearance, and microendoscopic imaging does not add a significant amount of time to a regular puncture biopsy, the path planning can be performed in a couple of minutes, while ICG is been cleared (15 minutes) from the plasma. Besides, the radiation from repetitive CT scans is drastically reduced because only two scans are needed. The use of this system may help to overcome the current limitations microendoscopy imaging must face in order to become clinically translatable: shallow penetration and emission signal in deep-seated tissues or organs. This is achieved by bringing the imaging tools closer to the target. Accuracy to reach the tumor also may increase when using the system, in our series the imaging of the tumor vessel were recorded mainly from the tumor center where a later biopsy will have more reliable result, this was validated by the presence of VX2 carcinoma cells in all the biopsies. The use of ICG takes advantages of the high angiogenesis rate seen in tumors, making the leaky vasculature [14] easy to visualize by microendoscopy imaging. Our results showed a difference in contrast between lung parenchyma and the region recognized as the VX2 tumor in CT scan where the contrast was higher whenever the fibered microendoscope was closer to the tumor core. This differentiation based on tumor angiogenesis also gives us different contrast level inside the tumor as we showed when comparing cases where the needle tip and fibered microendoscope had a peripheral location, a zone usually more fibrotic and therefore less vascular. This contrast differentiation could give an extra tool to the physician to determine a region for biopsy, first to distinguish normal tissue and then a zone more vascular from one less vascular.

Pharmacokinetics of ICG enables the potential to provide new tools for tumor detection, through the assessment of the leakiness of large ICG molecules from the microvasculature; this permeability is a characteristic of the poorly developed vasculature observed in tumor angiogenesis. The increase in local microvasculature density is also expected to induce a high perturbation in the optical signal from intercapillary vessel inside the tumor. Therefore, considering all the characteristics offered by the use of ICG in conjunction with microendoscopic imaging, this approach may add functional information to morphology, all in real time.

The clinical impact of this technology could be extremely relevant considering that biopsy is a sine qua non condition for tumor diagnosis. Due to tumors heterogeneity,

it can be difficult to establish the optimal site for tissue biopsy and sampling error may lead to incorrect diagnoses. The use of microendoscopy probes for tumor detection using ICG assisted by real time 3D CT image navigation is a technique that could improve the sensitivity of tumor detection, several factors contribute to this end: 1) Continuous tracking of the needle after it has been introduced to the chest without taking more CT scans, 2) noticeable visualization of tumor regions based on their vascularity, and 3) easy recognition of necrotic regions, capsule, and lung parenchyma from tumor (Fig. 4) as they represent non-vascular areas and therefore ICG cannot exert its diffusible (extravascular) flow in leaky capillaries. In our experiments, all the biopsies taken from the tumors were positive; these biopsies were taken from the regions were a higher ICG uptake was detected by microendoscopic imaging.

Among the limitations we faced in our study we have a relative small series of cases, use of VX2 line rather than an adenocarcinoma line, and lack of sensors attached to the microendoscope to extend the electromagnetic tracking capabilities to the microendoscope. The VX2 line was chosen based on their easiness of handling, survival rate, and time to growth. To address the last limitation our group has been working in adding EM-sensors to track the fiberoptic image probe to ease fusion of microendoscopy video with 3D lung CT volume, therefore a better correlation and localization among needle, microendoscope, and tumor vessel imaging can be obtained, all together and in real time.

A minimally invasive technique that provides in vivo optical expression in real-time would radically improve clinical diagnosis by offering a better assessment of changes in tissues. Microendoscopic imaging can meet this important need by delivering real-time, in vivo images with sub-cellular resolution, and the addition of optical-specific information to anatomic images may benefit the survival rate of patients with peripheral lung tumors by providing an earlier and more specific detection of the disease.

5 Conclusions

Our research findings indicates that using real-time 3D CT image guided intervention to guide a microendoscopic probe inside a tumor is fast and effective. This new approach could be extremely relevant in lung cancer detection where repeated rounds of CT scans are necessary both for screening, diagnosis and post-treatment follow-up. This system allows an accurate position of the needle and the microendoscope inside the tumor for in situ microendoscopic imaging. The addition of in situ microendoscopic imaging is especially useful during biopsy since it may allow a better characterization and guidance. Moreover, the addition of tumor vessels information to the tumor images provides by ICG has the potential to provide earlier and more specific tumor detection. These same features also can contribute to a more selective tumor biopsy or ablation. To prove this concept our group is designing a clinical trial in collaboration with interventional radiologists to assess the real time 3D CT Image Navigation System and microendoscopic imaging with ICG for peripheral lung tumor detection and treatment in selected patients.

References

1. Parkin, D.M., Bray, F., Ferlay, J., Pisani, P.: Global cancer statistics, 2002; CA Cancer J. Clin. 55(2), 74-108 (March-April 2005)
2. Cook, R.J., Paolinelis, G., Mannocci, F., Banerjee, A., Watson, T.F.: A new flexible confocal micro-endoscopic technology for in vivo diagnosis of soft and hard tissue intra-oral lesions (2005); In Focus on Microscopy
3. Yang, X.: Interventional molecular imaging. Radiology 254(3), 651–654 (2010)
4. Wong, K., Valdivia y Alvarado, M., He, T.C., Zhang, Y., Xue, Z., Wong, S.T.: Real time in situ molecular imaging in lungs tumors using image guided 3D computed tomography. In: CARS 2010, Geneva, Switzerland (2010)
5. Thiberville, L.C., Bourg-Heckly, G., Peltier, E., Cavé, C.: In vivo endoscopic analysis of the bronchial structure using fluorescence fibered confocal microscopy. In: European Respiratory Society, ERS 2006 (2006)
6. Cherrick, G.R., Stein, S.W., Leevy, C.M., Davidson, C.S.: Indocyanine green: observations on its physical properties, plasma decay, and hepatic extraction. J. Clin. Invest. 39, 592–600 (1960)
7. Alacam, B., Yazici, B., Intes, X., Chance, B.: Extended kalman filtering for the modeling and analysis of ICG pharmacokinetics in cancerous tumors using NIR optical methods. IEEE Trans. Biomed. Eng. 53(10), 1861–1871 (2006)
8. He, T.C., Xue, Z., Wong, K., Valdivia y Alvarado, M., Zhang, Y., Xie, W., Wong, S.T.: A Minimally Invasive Multimodality Image-Guided (MIMIG) Molecular Imaging System for Peripharal Lung Cancer Intervention and Diagnosis. In: Navab, N., Jannin, P. (eds.) IPCAI 2010. LNCS, vol. 6135, pp. 102–112. Springer, Heidelberg (2010)
9. Tu, M., Xu, L., Wei, X., Miao, Y.: How to establish a solitary and localized VX2 lung cancer rabbit model? A simple and effective intrapulmonary tumor implantation technique. J. Surg. Res. 154(2), 284–292 (2009); Epub 2008 (July 15, 2008)
10. Shomura, Y., Saito, Y., Minami, K., Imamura, H.: A new method for establishing an intrapulmonary tumor in the rabbit. Jpn. J. Thorac. Cardiovasc. Surg. 51(8), 337–343 (2003)
11. Thiberville, L., Salaun, M., Lachkar, S., Dominique, S., Moreno-Swirc, S., Vever-Bizet, C.: Human in-vivo fluorescence microimaging of the alveolar ducts and sacs during bronchoscopy. Eur. Respir. J. 33(5), 974–985 (2009)
12. Solomon, S.B., White Jr., 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(6), 1783–1787 (2000)
13. Solomon, S.B., White, P., Acker, D.E.: Real-time bronchoscope tip localization enables three-dimensional CT image guidance for transbronchial needle aspiration in swine. Chest 114, 1405–1410 (1998)
14. Chen, Y., Liu, Q., Huang, P., Hyman, S., Lee, W., Chance, B.: Assessment of tumor angiogenesis using fluorescence contrast agents. In: Proc. of the SPIE, vol. 5254, pp. 296–301 (2003)

Particle-Based Deformable Modeling with Pre-computed Surface Data in Real-Time Surgical Simulation

Bo Zhu, Lixu Gu, and Zhe Zhou

Image Guided Surgery and Therapy Lab, Shanghai Jiao Tong University, China
{boolzhu,gulixu,ziziamy}@sjtu.edu.cn

Abstract. Particle-based method has proven to be a powerful tool in real-time surgical simulation for its simplicity and high efficiency. However, it is difficult to model the high-resolution surface deformation details with standard particle-based techniques. In this paper, we propose a novel approach to model the elastic behaviors of organs with complex surfaces in surgical environment. The basic idea of our approach is to introduce an auxiliary surface mesh into the existing particle-based simulation framework, and utilize the pre-computed surface data with experimental biomechanics parameters for deformable modeling. The high-resolution organ deformations and the low-resolution soft tissue deformations are treated in surface-based and particle-based methods respectively. Our method provides an efficient and physical valid way to model the organ deformation details for particle-based surgery simulation techniques without using adaptive particle methods, as shown in our experiment results.

Keywords: interactive surgical simulation, minimally invasive surgery, deformable modeling, point-based simulation, physically-based modeling.

1 Introduction

Interactive surgery simulation techniques play an important role in today's medical training program. Modeling the deformable organs and tissues accurately in surgery scenes is a hot topic in real-time medical simulation. A variety of models has been proposed over the past decades to provide deformation results balancing accuracy and efficiency in real-time virtual surgery simulation (e.g. mass-spring model [13][14], finite element model [6], boundary element model [4][2] or particle-based model [5][7][9][10]).

Different data representations, such as volume mesh, surface mesh or particle systems, are used in different physics models. Because there exists hardly any direct correlation between these structures, to couple these models in one system with high performance is a challenge to real-time surgical simulations. Particle-based simulation methods have been proposed over the past decades to alleviate the simulation problems brought by meshes in the area of computation physics and computer graphics. For its simplicity in topology and data representation, particle-based methods are suitable for simulating the elastic and plastic bodies in real-time applications. And they have been introduced into minimally invasive surgery simulation in [8] and [10].

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 503–512, 2010.

However, there also remain several problems in pure particle-based methods. One of them is the simulation resolution. Since the objects in the scene are all represented by discrete points with the same support field, it is difficult to describe some deformation details on the organ surface, which is visually important in surgical training. Although some adaptive methods have been proposed in particle-based simulation in [1], they are less efficient in real-time simulation since extra data structures (e.g. octree) are required to handle the points with different resolutions. To model the deformation details without sacrificing the high efficiency, in this paper we propose an alternative method for the adaptive particle approach. In our approach, a surface mesh is re-introduced into the particle-based simulation framework and serves as a high-resolution elasticity solver with pre-computed surface deformation data. Unlike the previous mesh-based methods, the surface mesh in our methods is only used for physics computation, and all the other aspects of simulation, including collision detection, contact handling and haptic rendering, are treated in a particle-based way with the same resolution. A physically-based simulation framework with combined particle-based and surface-based physics representations is applied in simulating laparoscopic surgery training procedures. Insignificant elements in the surgical training scenes, such as small soft tissues and fats, are modeled with a standard particle-based method, while the important target pathologic organs are modeled with our new deformable model to provide high-resolution visual feedback to the trainees.

2 Overview

Our method is encouraged by the previous works of particle-based surgical simulation [8][9][10]. It contains three main parts outlined below.

1. **Point and Mesh Construction from Medical Data.** In this phase the medical data is segmented into different parts representing different organs, bones and tissues. A set of physics points is extracted and surface meshes are constructed for highly detailed simulation of pathologic organs and tumors. Distance field information is constructed based on the particles for fast collision handling.
2. **Surface Data Pre-computing for Pathologic Organ Modeling.** For the important pathologic organs in surgical training, a deformation surface data is pre-computed using the boundary element elastic solver. The biological material properties measured by biomechanics experiments and the surface geometry information are used in the surface data pre-computation.
3. **Real-time Deformable Modeling with Hybrid Methods.** In runtime simulation, the pre-computed surface data is integrated into the particle-based simulation framework and help to guide the high-resolution organ deformation in a physically valid way. The high-resolution pathologic organs and other ordinary elements in the surgery environment are modeled in different ways.

3 Physics Point Extraction and Distance Field Construction from Medical Data

In this stage, organs and tissue models in the surgery environment are segmented from the CT image data. The discrete physics points are extracted from each segmented organ.

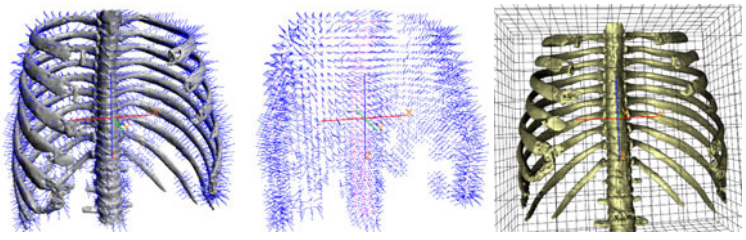


Fig. 1. In the data extraction stage, the distance field (left and middle) are constructed and put into the unified Eulerian grid for collision handling

For the low resolution deformable models, these points are directly used in simulation. For the high resolution models, an additional surface mesh is used to represent the surface of deformable object. The surface meshes are generated from medical data by using marching cube algorithm, and the particle surface is generated by sampling the coarse triangle mesh.

All the particles, including the particle surface used for high-resolution models and physics particles used for low-resolution models, are mapped onto a global grid for collision handling. Additionally, as illustrated in Fig. 1, we construct a local Eulerian grid for each immobile or rigid object such as bones and surgical instruments in the surgery scene, and pre-compute a distance field in the local Eulerian grid as in [1], [3] and [11]. The distance field is used in spatial-hashing based collision detection [15] for fast collision detection.

4 Particle-Based Fast Modeling for Small Soft Tissues

The target pathologic organs and the small soft tissues in surgical training are modeled in different ways in our framework. With the extracted physics points in the pre-processing step, we use smoothed particle hydrodynamics (SPH) to simulate the unimportant elements such as small soft tissues and fats in the scene. The deformable model is assumed as a Hookean material. The strain ε and the stress σ follow the following rules:

$$\sigma = \mathbf{C}\varepsilon \tag{1}$$

in which **C** is a rank four tensor. As for isotropic materials, **C** can be represented by a 6x6 matrix determined by Young's Modulus and Poisson Ratio of elastic materials as in [7]. In each time step, the surface force, body force and elastic force on each particle are summed up and applied in a common ODE solver to model the deformation. As in Fig. 2, different operations can be exerted on the fast point-based model to change the shapes of the object in a low resolution. We employ this pure particle-based method to model the small and unimportant soft tissues. But for the pathologic organs which usually play a central role in surgical training, a model with higher resolution is needed to model the organ deformation details in the interactions with the surgical instruments.

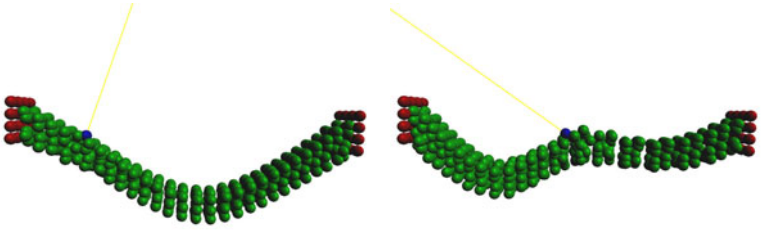


Fig. 2. Low-resolution deformations of the small soft tissues are modeled by using the pure particle-based method in our framework

5 Surface-Guided Modeling for Surgical Organs

To model the pathologic organs in the surgery scene precisely, surface data is pre-computed before runtime simulation and is used to provide accurate global deformation information for the particles in real-time deformable modeling. The surface data is a global deformation matrix calculated from the geometry and the material properties of the target surgical organ. It contains the deformation data of all the surface elements with higher resolution than the physics particles and can be used to guide the deformation of the particles in the runtime simulation.

The surface data pre-computing is based on the Boundary element method (BEM), but different from the previous BEM work [4][2][12], we only utilize the surface data (the global deformation matrix) pre-computed by BEM to guide the particle deformation, rather than directly computing the deformation with BEM in each timestep. All the computations related to BEM happen in the pre-computation stage instead of real-time loops. So there's no need to solve the boundary integration equations in each timestep as the standard BEM does. This process is replaced by setting external forces from surgical instruments and internal repulse forces from constraint particles into a boundary traction vector and calculating the global deformation by simply multiplying this vector with the pre-computed deformation matrix. Then the deformation computed from the surface data is regarded as a corrected target position for the physics particles. It backwardly corrects the displacement of these discrete particles and interacting with other particles in a Lagrangian way. The hybrid deformable model consists of two main parts: the surface data computation in the pre-computing stage and the surface guided particle deformation in real-time simulation.

5.1 Surface Data Pre-computation

In the surface data pre-computing stage, material parameters for a surface model are determined by biomechanics experiments, and the global deformation matrix of surface elements is initialized by using the geometry information of the coarse surface mesh and the measured biology material parameters. Three material parameters are used in the surface data pre-computing: Young's Modulus E (100-110kPa), Poisson's ratio σ (0.46-0.50) and density(1.00-1.12gcm-3). The former two parameters are used in surface data precomputing and are measured by determining the biomechanics

strain-stress relationships of tissues sampled from different parts of porcine kidney and liver in INSTRON mechanical testing machine as in [10].

With the biological material parameters and the surface geometry shape of the organ, the deformable surface data is calculated based on the Navier's equation of isotropic elastic material. The Green-Gauss theorem and the Kelvin fundamental solutions of linear elastostatic problem is used to set up an integral equation defined on the boundary of domain Ω and numerically discretized into a matrix formulation as

$$\mathbf{C}\mathbf{u}_i + \sum_{j=1}^{N}\left(\int_{\Gamma_j}\mathbf{p}^* d\Gamma\right)\mathbf{u}_j = \sum_{j=1}^{N}\left(\int_{\Gamma_j}\mathbf{u}^* d\Gamma\right)\mathbf{p}_j \quad (i=1,\cdots,n) \tag{2}$$

Take the summary of left and right side as H and G. To compute the integral term in $\mathbf{H}_{ij}$ and $\mathbf{G}_{ij}$, Gauss integration method is employed on surface element i and j. Substituting Kelvin fundamental solutions into Gauss integration formula, we get

$$\mathbf{H}_{ij} = |S_j|\,\omega_k \sum_{k=1}^{7}\mathbf{u}_{ij}(P,Q_k) \tag{3}$$

$$\mathbf{G}_{ij} = |S_j|\,\omega_k \sum_{k=1}^{7}\mathbf{p}_{ij}(P,Q_k) \tag{4}$$

where S_j is the area of element j, ω_k are Gauss coefficients and P, Q are sampled points on the triangle surface. In our implementation we use the seven point integration method for each triangle element. Then we get the linear equation system

$$\sum_{j=1}^{N}\mathbf{H}_{ij}\mathbf{u}_j = \sum_{j=1}^{N}\mathbf{G}_{ij}\mathbf{p}_j \quad (i=1,\cdots,N) \tag{5}$$

H and **G** are 3N×3N matrices, each containing N×N sub-matrices $\mathbf{H}_{ij}$ and $\mathbf{G}_{ij}$. The values in each sub-matrix are calculated based on the material parameters and the relative positions of the two triangle elements as in the standard BEM [4][12]. **U** and **P** are 1×3N vectors composed of displacement and traction vectors of N surface elements. Then the global deformation U is computed as

$$\mathbf{U} = \mathbf{H}^{-1}\mathbf{G}\mathbf{P} = \mathbf{M}\mathbf{P} \tag{6}$$

where **M** is the global deformation matrix. Computing **M** is a time exhausting task since it is a non-sparse and asymmetric matrix. But this surface data needs to be computed only once in the preprocessing stage and can be used to guide the surface deformation in runtime simulation.

5.2 Runtime Organ Deformation Simulation

In runtime simulation, the physically accurate global deformations are computed based on the global deformation matrix **M** according to the current boundary condition (by multiplying the matrix M with the surface traction vector), and then

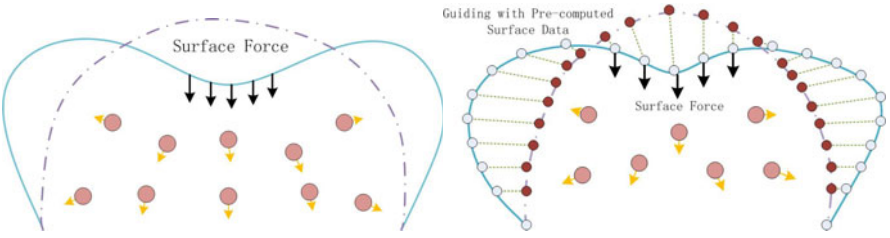


Fig. 3. Comparison between particle-based deformation (left) and our surface-guiding (right) method. High-resolution deformations are modeled with the pre-computed surface data.

served as target positions for the particles on the surface. As illustrated in Fig. 3(right), in each time step, displacements of physics particles are corrected based on the well-defined goal positions by virtual springs. They are used to describe the organ dynamic temporal deformations such as viscosity and creeping in runtime simulation. The values are determined by matching the continuous images taken from dynamic deforming organs.

The interactions between the deformable model and VR environment, including collisions, contacts and dynamic simulations, are handled by the particle surface using point-based techniques. So there is no interaction between the fine surface elements and the coarse physics particles with different support radius. All the computations are based on physics particles in the point-based framework, while the precomputed matrix **M** is the only extra data used in real-time simulation.

With the precomputed deformation matrix **M**, it is easy to compute the accurate deformation position for each physics particle in real-time without employing any adaptive particle method. As illustrated in Fig. 4, the local deformation of the pathologic liver in response to the surgical operations on it is modeled easily with the pre-computed matrix, which cannot be resolved with the low number of physics points (less than 1,000 in this case).

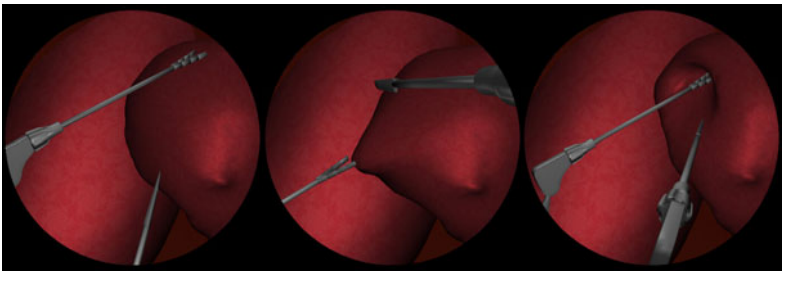


Fig. 4. The surface-based deformable model is integrated into our particle-based surgical minimally invasive surgical simulator. Deformation details of the liver organ in response to the surgical operations are modeled in real-time.

6 Experiments and Evaluations

Our particle-based deformable model with pre-computed surface data is integrated into a virtual reality training system for laparoscopic surgery as in Fig. 5. The training system is set up based on particle-based simulation techniques. Surgeons are trained to perform operations on pathologic organs with different biological parameters and acquire specific skills under different conditions. The simulation framework is implemented in C++ and rendered using OpenGL. All the experiments were carried out on a 2.26GHz Pentium M notebook PC with GeForce 9650M graphics card and 2 GB of memory.

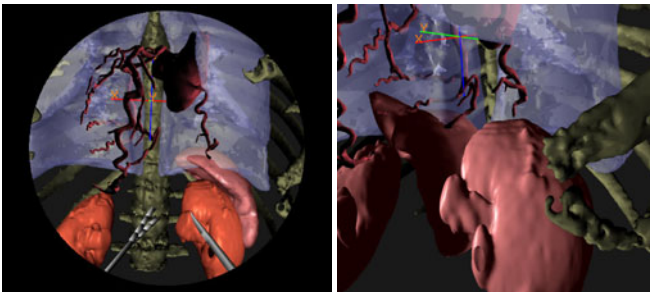


Fig. 5. Our virtual reality laparoscopic surgery training system employing the particle-based techniques and our novel model for pathologic organ deformations.

To evaluate the time performance of the surface-guided deformable model, the simulation method is tested on five models with different surface polygon numbers ranging from 600 to 2200. As in Table 1, summing the computation time of surface deformation, position correction and surface interpolation for each model, it can be seen from the overall time cost of each object that our hybrid model is no more than 5 ms and is suitable for real-time applications. Comparing with the standard particle-based method in [10], which can provide 60 fps simulation results, our particle-based method utilizing the pre-computed surface data can work within the same time restrictions (no less than 50 fps).

In the second experiment, we compare the accuracy of our method with the standard BEM. In our hybrid model, the deformation of the surface is computed by

Table 1. Time performance of the surface-based deformable model

Model	Point Number	Surface Computation(ms)	Position Correction(ms)	Surface Interpolation(ms)	Overall (ms)
1	284	0.63	0.228	0.137	1.00
2	508	1.13	0.309	0.266	1.67
3	784	1.75	0.653	0.422	2.83
4	876	1.94	0.725	0.497	3.162
5	1144	2.53	1.012	0.634	4.176

utilizing a pre-computed surface data, and interpolated back to the discrete particles. Compared with traditional surface-method such as BEM, it is unavoidable that some deformation details may be lost during interpolation between surface and particle models. We measure this accuracy loss by comparing our method with the standard BEM method as in Table 2. The accuracy loss ratio R is defined as $(\sum \|\mathbf{u}_i^0 - \mathbf{u}_i^{intp}\|) / (\sum \|\mathbf{u}_i^0\|)$, in which $\mathbf{u}_i^0$ is the vertex displacements computed by using standard BEM model, and $\mathbf{u}_i^{intp}$ is vertex displacements computed using surface interpolation.

Table 2. Relationship between vertex number ratio and accuracy loss ratio

Vertex Number (Coarse)	Vertex Number (Fine)	Polygon Number (Fine)	Vertex Num Ratio	Accuracy Loss Ratio
284	256	508	0.90	2.1%
	394	784	1.39	3.7%
	574	1144	2.02	5.1%
608	506	1008	0.83	1.4%
	1000	1996	1.64	3.3%
	1586	*3168*	*2.61*	6.2%

It is concluded from Table 2 that for most organ models in surgical simulation with relatively smooth surface, this accuracy loss can be restricted within an acceptable scale if the ratio of coarse surface vertex number to fine vertex polygon number is below a threshold. When the vertex number ratio is below 2, the accuracy loss ratio can be restricted below 5%-6%, which is not obvious in visualization. Considering that the time complexity of pre-computing global deformation matrix is $O(n^2)$, and updating global matrix in real-time is $O(n)$, it is worthwhile to accelerate computing time 4-5 times in the cost of tiny visualization lost.

Compared with the pure particle-based method, our hybrid model can produce high-resolution local surface details in response to surgical instruments as in Fig. 4, and these details cannot be resolved on the physical particle level. There are also some limitations of our approach. The temporal elastic behavior based on spring parameters is less accurate than those based on continuous mechanics. The main purpose of our approach is to simulate the high-resolution organ surface deformation as accurately as possible (by using precomputed data) and in other aspects to provide trainees visually plausible feedbacks. Future work includes designing more accurate biomechanical experiments to measure the real material parameters and applying the method to more complex and heterogeneous models.

7 Conclusion

In this paper we propose a new method to model high-resolution organ deformations within particle-based simulation framework. We utilize the precomputed surface data

with experimental biological material parameters to model the deformation details on the surface of important organs in a physically-based way. Our method provides a new alternative for multi-resolution surgical simulation with particle-based methods. Since most of the material deformation details are precomputed and stored, our method works well in real-time virtual surgery environment and provides physically valid simulation results.

Acknowledgement

This work is partially supported by the Chinese National 863 Research Foundation under Grand No.2007AA01Z312, the Chinese National Natural Science Foundation under Grant No.60872103 and the Med-X Research Funding of SJTU under Grand No.YG2009ZD101.

References

1. Adams, B., Pauly, M., Keiser, R., Guibas, L.J.: Adaptively sampled particle fluids. In: ACM SIGGRAPH 2007, p. 48. ACM, New York (2007)
2. Zhu, B., Gu, L., Zhang, J., Yan, Z., Pan, L., Zhao, Q.: Simulation of organ deformation using boundary element method and meshless shape matching. In: EMBC 2008, IEEE Engineering in Medicine and Biology Society, pp. 3253–3256 (2008)
3. Harada, T., Koshizuka, S.: Real-time cloth simulation interacting with deforming high-resolution models. In: ACM SIGGRAPH 2006 Research posters, p. 129. ACM, New York (2006)
4. James, D.L., Pai, D.K.: ArtDefo: accurate real time deformable objects. In: ACM SIGGRAPH 1999 papers, pp. 65–72 (1999)
5. Müller, M., Keiser, R., Nealen, A., Pauly, M., Gross, M., Alexa, M.: Point based animation of elastic, plastic and melting objects. In: SCA 2004: Proceedings of the 2004 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, pp. 141–151 (2004)
6. Müller, M., Teschner, M., Gross, M.: Physically-Based Simulation of Objects Represented by Surface Meshes. In: Proc. of Computer Graphics International, pp. 156–165 (2004)
7. Solenthaler, B., Schläfli, J., Pajarola, R.: A unified particle model for fluid–solid interactions: Research Articles. Computer. Animation Virtual Worlds 18(1), 69–82 (2007)
8. Turini, G., Pietroni, N., Megali, G., Ganovelli, F., Pietrabissa, A., Mosca, F.: New techniques for computer-based simulation in surgical training. International Journal of Biomedical Engineering and Technology, IJBET (2008)
9. Wang, X., Chen, T., Zhang, S., Metaxas, D., Axel, L.: LV Motion and Strain Computation from tMRI Based on Meshless Deformable Models. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 636–644. Springer, Heidelberg (2008)
10. Zhu, B., Gu, L., Peng, X., Zhou, Z.: A Point-Based Simulation Framework for Minimally Invasive Surgery. In: Bello, F., Cotin, S. (eds.) ISBMS 2010. LNCS, vol. 5958, pp. 130–138. Springer, Heidelberg (2010)
11. Francois, F., Jeremie, A., Matthieu, N.: Eulerian Contact for Versatile Collision Processing. INRIA technical report RR-6203 (2007)

12. Kim, J., Choi, C., De, S., Srinivasan, M.A.: Virtual surgery simulation for medical training using multi-resolution organ models. The International Journal of Medical Robotics and Computer Assisted Surgery 3(2), 149–158 (2007)
13. Teschner, M., Heidelberger, B., Muller, M., Gross, M.: A Versatile and Robust Model for Geometrically Complex Deformable Solids. In: CGI 2004: Proc. of the Computer Graphics International, pp. 312–319 (2004)
14. Bianchi, G., Solenthaler, B., Szekely, G., Harders, M.: Simultaneous Topology and Stiffness Identification for Mass-Spring Models Based on FEM Reference Deformations. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3217, pp. 293–301. Springer, Heidelberg (2004)
15. Teschner, M., Heidelberger, B., Mueller, M., Pomeranets, D., Gross, M.: Optimized spatial hashing for collision detection of deformable objects, pp. 47–54 (2003)

Direct Co-calibration of Endobronchial Ultrasound and Video

Philipp Dressel[1], Marco Feuerstein[1,2], Tobias Reichl[1], Takayuki Kitasaka[2,3], Nassir Navab[1], and Kensaku Mori[2]

[1] Technische Universität München
[2] Nagoya University
[3] Aichi Institute of Technology

Abstract. This paper for the first time presents a fast and practical approach for the co-calibration of endobronchial ultrasound (EBUS) and video images from a single image pair. At comparable accuracy, it achieves a significantly higher precision than a reference method based on combining electromagnetic tracking with two standard calibration techniques (hand-eye and single-wall calibration). Its resulting calibration matrix can be utilized in navigated bronchoscopy to allow easier EBUS manipulation and positioning.

1 Introduction

Endobronchial ultrasound (EBUS) recently became a valuable source of information in transbronchial interventions, particularly for verification of lesion position and ultrasound-guided biopsies that were previously performed blindly or under fluoroscopy exposing both patient and staff to ionizing radiation. However, smooth insertion and manipulation of the EBUS bronchoscope to obtain clear images is still difficult due to its forward oblique view [1].

At the same time, navigated bronchoscopy is increasingly used in order to facilitate instrument manipulation and increase operator confidence by visualizing the bronchoscope pose in relation to a preinterventional computed tomography (CT) image [2]. Navigation can be accomplished, if the bronchoscope is localized e.g. by external electromagnetic tracking [3], by registering bronchoscope images to virtual images created from CT [4,5], or by hybrid methods combining these two approaches [6]. However, navigated bronchoscopy and its advantages have not yet been transferred to EBUS.

This paper provides the first step for achieving this transfer. In order to spatially relate images acquired by the ultrasound transducer and the camera integrated into the bronchoscope, we propose a fast and practical approach for computation of the Euclidean transformation between their associated coordinate systems. After the camera is localized by electromagnetic tracking or image registration, we can automatically determine the position and orientation of both the camera image and the ultrasound plane in relation to a preinterventional CT image, enabling navigation and hence easier EBUS insertion and positioning.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 513–520, 2010.

2 Method

We propose a two-step method to compute the spatial relation ${}^{C}\boldsymbol{T}_{U}$ between camera image and ultrasound plane of an EBUS system as shown in Fig. 1). The first calibration step consists of measuring the geometry of a custom calibration phantom using a calibrated pointer and a tracking system (refer to Fig. 2). The phantom consists of an optical pattern used to estimate the bronchoscope camera pose and a Z-fiducial configuration of rubber tubes to estimate the ultrasound pose. In the second calibration step, the transformation ${}^{C}\boldsymbol{T}_{U}$ can be computed from a single pair of camera and ultrasound images.

The measurement of the phantom geometry needs to be performed only once, and a tracking system is not required for subsequent EBUS calibrations. If the phantom is precision manufactured, the first calibration step may be omitted entirely.

The mapping between ultrasound image coordinates (u, v) and camera image coordinates (x, y) is goverened by the following equation:

$$\lambda \cdot \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \left[K \middle| 0 \right] \cdot {}^{C}\boldsymbol{T}_{U} \cdot \begin{pmatrix} s_x \cdot u \\ s_y \cdot v \\ 0 \\ 1 \end{pmatrix} \tag{1}$$

The ultrasound pixel scaling (s_x, s_y) and the subsequent transformation ${}^{C}\boldsymbol{T}_{U}$ from the ultrasound plane into camera coordinates are followed by the projection

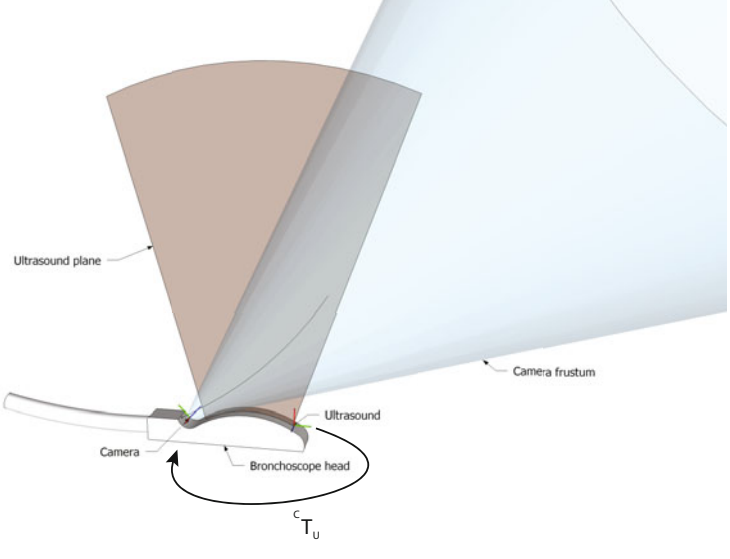


Fig. 1. EBUS system. The ultrasound plane and camera frustum in their approximate spatial relation.

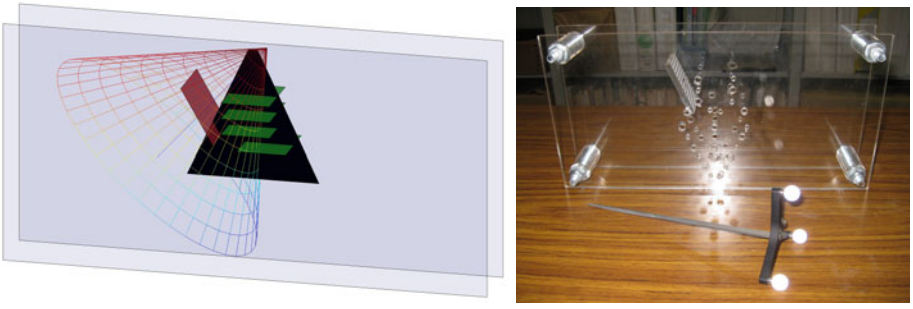


Fig. 2. Left: model of the calibration phantom. The cone depicts the camera frustum, the trapezoid represents the ultrasound plane. Planes mark the positions of Z-fiducial layers, optical pattern and supporting structure, respectively. Right: phantom during measurement with a tracked pointer. The tubes of the Z-fiducials have been removed for easier measurements.

through the camera matrix K and, not shown in Eq. 1, by radial and tangential distortion. As K and the calibration coefficients are intrinsic to the camera and hence do not change, they can be precalibrated using a standard camera calibration method with a standard printed pattern in air, e.g. the one proposed by Wengert *et al.* [7]. This allows for optimal freedom in camera/pattern positioning, as opposed to working within the constraints of a custom phantom. However, during ${}^{C}\boldsymbol{T}_{U}$ calibration the camera will be placed underwater to allow for simultaneus acquisition of camera and ultrasound images. The effects of this submersion need to be accounted for during the computation of the transformation matrix.

2.1 Underwater Camera

The submersion of the camera during calibration causes a magnification of the recorded images, affecting both the intrinsic camera parameters and distortion coefficients modeling the camera system. According to Lavest *et al.* [8], the focal length of the camera model needs to be adjusted by the ratio of the optical indices of the involved media. For an air to water adaptation, this ratio is $r = \frac{4}{3}$. However, the distortion coefficients cannot be adjusted without a detailed model of the camera's lens system. To overcome this, we instead apply the ratio r as a magnification directly on the image plane, after any projections and distortions are performed using the camera parameters calibrated in air. Lavest *et al.* empirically prove this method to be valid [8]. Any point p_a on the image plane in air can thus be mapped to coordinates p_w in water with the following equation:

$$p_w = \alpha + (r \cdot (p_a - \alpha)) \tag{2}$$

where α is the camera's principal point.

2.2 Phantom Design

To calibrate the transformation ${}^{C}\boldsymbol{T}_{U}$ from a single pair of camera and ultrasound images the phantom needs to allow for simultaneous estimation of camera and ultrasound poses. To estimate the camera pose, a dot pattern is printed on a decal with an office laser printer. This decal is attached to an acrylic glass plate. The dot pattern layout used in the automatic calibration toolbox implementing the method proposed by Wengert *et al.* [7] is used with a scaled dot spacing of 5 mm.

To estimate the ultrasound pose, a Z-fiducial setup based on the method of Comeau *et al.* [9] is used. Four layers of Z-fiducials are used, each containing one Z-shape comprised of a rubber tube filled with water. The Z-fiducials are supported by two acryic glass panels, spaced 40 mm apart. The plate containing the optical pattern is also glued to this supporting structure. Figure 2 shows the setup of the phantom.

2.3 Calibration

For convenience, the origin of the phantom coordinate system is defined to coincide with the origin of the optical pattern coordinates. During the first calibration step, the location of the Z-fiducials is computed in phantom coordinates. For this, the location of the supporting holes as well as the location of the optical pattern are measured.

During ${}^{C}\boldsymbol{T}_{U}$ calibration the computed poses ${}^{P}\boldsymbol{T}_{C}$, ${}^{P}\boldsymbol{T}_{U}$ of camera and ultrasound are thus within the same coordinate system, and the desired transformation can be computed trivially as: ${}^{C}\boldsymbol{T}_{U} = {}^{P}\boldsymbol{T}_{C}{}^{-1} \cdot {}^{P}\boldsymbol{T}_{U}$.

The user inferface for the calibration process is shown in Fig. 4.

Camera Pose Estimation. The feature detection for the camera image is based on the algorithm proposed by Wengert [7]. The pattern is used unmodified (except for scaling) and the feature matching algorithm is analogous to the originally proposed version. The preprocessing and blob extraction before the feature matching was adapted to the underwater, real-time environment of the calibration procedure.

Figure 3 (right) outlines the camera pose estimation algorith. The color video image from the camera is first converted to grayscale. It is then subjected to a histogram equalization, followed by an (inverse) adaptive thresholding to produce a binary image containing the dots and bars of the pattern. The adaptive threshold allows for a robust extraction of the features even under the usually very inhomogeneous illumination caused by the bronchoscope's light source. Blobs are then extracted from the binary image. Before being processed, the detected points are scaled as if they were in an air environment, so that the intrinsic parameters calibrated in air can be used for the pose estimation. The two large bars defining the local coordinate system of the pattern can be identified through their large ellipticity, defined as the ratio of their major and minor axis length. For the bars, this ratio is greater than two, while the circles have a value

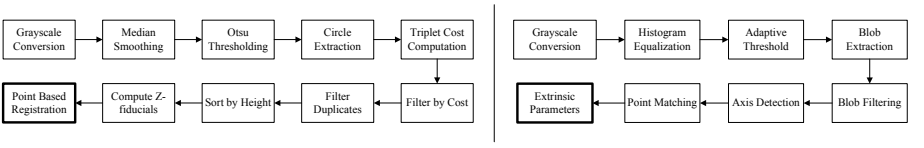


Fig. 3. Ultrasound pose (left) and camera pose (right) estimation steps

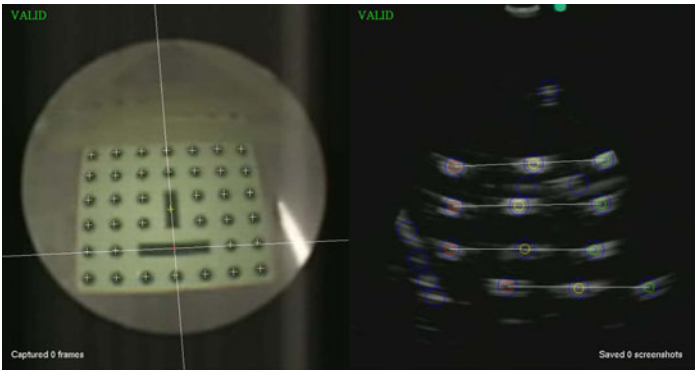


Fig. 4. Screenshot of the user interface during calibration

of roughly one. The two bars are then distinguished by their area. Once the bars are detected, an iterative search is performed, matching the circles of the pattern to its local coordinate system. Once the point correspondences are established, a projective transformation between all points is computed using the OpenCV library[1].

Ultrasound Pose Estimation. Figure 3 (left) shows the ultrasound pose estimation algorithm. The ultrasound image is first converted to grayscale, smoothed with a median filter and subjected to a binary thresholding. From the resulting binary image, all contours are extracted using OpenCV. The extracted contours are then filtered based on an area threshold to eliminate small artifacts. From the resulting list of candidates, every possible triplet is assembled and associated with a cost. The cost function is defined as the sum of squared distances of the points on the y-axis of the image. For any given triplet of y-coordinates c_1, c_2, c_3, the cost is thus:

$$C(c_1, c_2, c_3) = (c_1 - c_2)^2 + (c_1 - c_3)^2 + (c_2 - c_3)^2 \tag{3}$$

where c_i are the y-coordinates of the image points in the triplet. Points are then filtered with a cost threshold. No detected contour can be part of two fiducials, but all points are part of multiple triplets since all possible permutations are

[1] http://sourceforge.net/projects/opencvlibrary/

considered. To correct this, the triplets are iterated through, starting from the lowest cost triplet. Any subsequent triplet containing one of the points of the current triplet is discarded, ensuring all triplets contain unique points. Finally, the remaining triplets are sorted by their mean y-position, and the points of the triplets are sorted by their x-position. The n triplets with the lowest cost now contain, in order, the points of the n Z-fiducials. Using the method of Comeau *et al.* [9], n points in 3D space can be computed from the detected fiducials. These points are registered to the phantom geometry with a point-based registration, yielding the desired pose of the ultrasound probe.

3 Experiments

To establish a reference transformation ${}^{C}\boldsymbol{T}_{U}$ we attached a 6DOF position sensor of an NDI Aurora tracking system to the EBUS system. Using the hand-eye calibration method proposed by Tsai and Lenz [10], we compute the transformation ${}^{S}\boldsymbol{T}_{C}$ between camera image coordinates and the sensor's local coordinate system. Additionally, we compute the transformation ${}^{S}\boldsymbol{T}_{U}$ from the ultrasound image plane to the sensor using the single-wall method proposed by Prager *et al.* [11].

For a series of five hand-eye calibrations and five single-wall calibrations, all 25 possible permutations for a resulting calibration transformation ${}^{C}\boldsymbol{T}_{U}$ are compared to a series of 25 transformations obtained with our proposed method.

To provide a quantitative measure of the calibration error, the tip of a probe tool was moved to intersect the ultrasound plane while at the same time being visible in the camera image. The resulting small dot in the ultrasound image provides coordinates (u, v) which can then be mapped into the camera image using the calibrated transformation and Eq. 1. The distance of this backprojected, distorted and scaled point to the actual tip location in the 320 by 320 pixel image can then be measured, giving a quantitative indication of the calibration quality.

Due to the thickness of the ultrasound plane, errors in positioning the tip on the plane can contribute significantly to the resulting error value. Since the ultrasound plane is nearly orthogonal to the camera image plane, effectively forming a vertical line in the camera image, only the difference in image height is considered for the error. This ensures that the error introduced by the verification method itself is minimal.

This pixel error value was computed for all 50 calibrations with the resulting cumulative error values for a series of 20 points being shown in Fig. 5.

Table 1. Statistics for the mean pixel error values for the 20 backprojected point pairs using the hand-eye- and single-wall-based (he+sw) and phantom-based calibrations

	phantom	he+sw
rms	6.75 px	32.8 px
std	0.44 px	23.88 px
min	6.11 px	5.97 px
max	7.82 px	83.88 px

Table 1 shows the root mean square (RMS), standard deviation (STD), minimum, and maximum of the cumulative error values for both series of calibrations. It is apparent that while the best calibra-

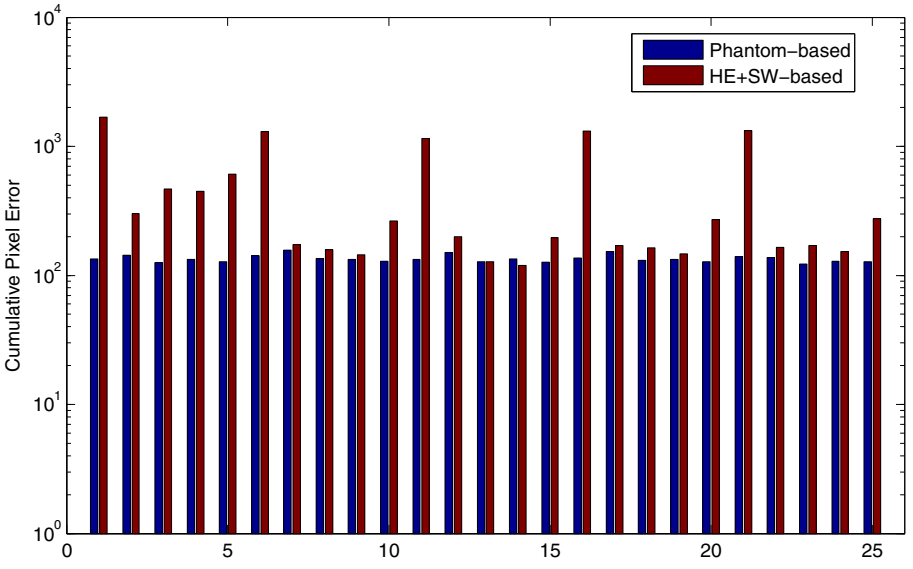


Fig. 5. Estimated cumulative errors for 20 point pairs for 25 hand-eye and single-wall (HE+SW) calibrations and 25 phantom-based calibrations. Please note that the y-axis has exponential scale.

tions of both series (i.e. the ones with the lowest error) are competitive, the hand-eye- and single-wall-based approach is less precise.

4 Discussion

As we can see from the results of our experiments, our proposed method is significantly more precise than a combination of electromagnetic tracking, hand-eye calibration, and single-wall calibration at a comparable maximum accuracy. We partially attribute this to the use of optical tracking during measurement of the phantom geometry, which has a higher precision and accuracy than electromagnetic tracking. Once a phantom is built or measured, our method can also be performed faster than hand-eye calibration and single-wall calibration, which is advantageous if several calibrations are required. Furthermore, the actual calibration does not require an external tracking system, which may not be essential in a laboratory setup, but can be very important for clinical or commercial applications.

After calibration, the transformation matrix ${}^{C}\boldsymbol{T}_{U}$ normally does not change any more. So even if external tracking is applied for navigated bronchoscopy, after attaching e.g. an electromagnetic tracking sensor, we only need to determine the transformation from sensor coordinates to camera coordinates without the need for ultrasound calibration in a water bath.

For further studies we are planning to use a precision machined phantom with known geometry to achieve an even higher precision and accuracy than with

optical tracking. This kind of phantom renders a tracking system unnecessary and can hence further reduce costs and complexity of our procedure. For our next generation phantom we will also try to improve the material properties of the Z-phantom tubes for better visibility in ultrasound images. Moreover, we are about to develop a similar calibration approach for endoscopes with a radial ultrasound transducer that have non-overlapping fields of view, along with the investigation of endobronchial 3D ultrasound.

5 Conclusion

This paper for the first time presents a fast and practical method and design for the co-calibration of endobronchial ultrasound and camera frames, which only requires a single image pair to perform calibration. At comparable accuracy, it achieves a significantly higher precision than a reference method based on electromagnetic tracking, hand-eye calibration, and single-wall calibration.

References

1. Herth, F.J.F., Eberhardt, R., Vilmann, P., et al.: Real-time endobronchial ultrasound guided transbronchial needle aspiration for sampling mediastinal lymph nodes. Thorax 61, 795–798 (2006)
2. Eberhardt, R., Anantham, D., Herth, F., et al.: Electromagnetic navigation diagnostic bronchoscopy in peripheral lung lesions. Chest 131, 1800–1805 (2007)
3. Wegner, I., Biederer, J., Tetzlaff, R., et al.: Evaluation and extension of a navigation system for bronchoscopy inside human lungs. In: SPIE Medical Imaging (2007)
4. Rai, L., Helferty, J.P., Higgins, W.E.: Combined video tracking and image-video registration for continuous bronchoscopic guidance. Int. J. CARS 3, 315–329 (2008)
5. Deguchi, D., Mori, K., Feuerstein, M., et al.: Selective image similarity measure for bronchoscope tracking based on image registration. Medical Image Analysis 13(4), 621–633 (2009)
6. Soper, T.D., Haynor, D.R., Glenny, R.W., Seibel, E.J.: In vivo validation of a hybrid tracking system for navigation of an ultrathin bronchoscope within peripheral airways. IEEE Trans. Biomed. Eng. 57(3), 736–745 (2010)
7. Wengert, C., Reeff, M., Cattin, P., et al.: Fully automatic endoscope calibration for intraoperative use. Bildverarbeitung für die Medizin, 419–423 (2006)
8. Lavest, J., Rives, G., Lapreste, J.: Underwater camera calibration. In: Vernon, D. (ed.) ECCV 2000. LNCS, vol. 1843, pp. 654–668. Springer, Heidelberg (2000)
9. Comeau, R., Fenster, A., Peters, T.: Integrated MR and ultrasound imaging for improved image guidance in neurosurgery. In: Proceedings of SPIE (1998)
10. Tsai, R., Lenz, R.: A new technique for fully autonomous and efficient 3D roboticshand/eye calibration. IEEE Transactions on Robotics and Automation 5, 345–358 (1989)
11. Prager, R., Rohling, R., Gee, A., et al.: Rapid calibration for 3-D freehand ultrasound. Ultrasound in Medicine & Biology 24, 855–869 (1998)

Real-Time Epicardial Excitation Time Map Overlay

Paul de Lange[1], Yuhei Takata[2], Hongho Kim[3], Hongen Liao[1], Etsuko Kobayashi[3], Minoru Ono[4], Shunei Kyo[4], Shinichi Takamoto[4], Satoshi Ishii[5], Takeo Asano[5], and Ichiro Sakuma[3]

[1] Bioengineering Department, The University of Tokyo, Japan
[2] Department of Frontier Sciences, The University of Tokyo, Japan
[3] Department of Precision Engineering, The University of Tokyo, Japan
[4] Department of Cardiothoracic Surgery, The University of Tokyo Hospital, Japan
[5] R&D Planning Division at Olympus Coporation, Japan

Abstract. This paper presents a method to display Epicardial Excitation Time (ECET) map information intuitively to a surgeon. The method uses Conditional Density Propagation to track the epicardium during the complete cardiac cycle. Then, combining this algorithm with electrophysiological mapping techniques, pre-recorded ECET maps are overlay onto the surgeons endoscope stream to give direct feedback on cardiac excitation propagation. To the best of our knowledge, this is the first report of electrophysiologic data being augumented with a surgeons endoscope stream.

1 Introduction

Arrhythmia is an irregularity in heart rhythm resulting from a defect in the conduction system of the heart [1]. The most common intermittent arrhythmia is *atrial fibrillation* (AF). AF leads to irregular impulses of the ventricles that generate the heartbeat and therefore results in an irregular heartbeat lasting from a few minutes to weeks. AF tends to become a chronic condition and increases the risk of death [2] [3].

Cardiac electrophysiologic information is used to locate abnormal conduction pathways. Typically, spatiotemporal endocardial potential is recorded and used to generate some cardiac excitation time map. This map must then be communicated to the electrophysiologist. There are several commercial systems that perform this step ("Carto System" by Biosence-Webster Inc., "Ensite catheter System" by St. Jude Medical co. Ltd., etc.). These systems always map the *endocardial* excitation via catheter insertion. However, we have been working with a group of thoracic surgeons who attempt to perform ablation on the *epicardium*. This is basically a variant of the Maze procedure [4]. To achieve this they require a system that maps excitation time of the epicardium. This paper will hereby document our solution to intuitive real-time EpiCardial Excitation Time (ECET) map visualization.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 521–530, 2010.

We attempt to display ECET map information directly onto the surgeons endoscope stream. However, in doing so we must deal with the problem of tracking the epicardium in real-time. Several groups are working in this area [5] [6] [7]. In particular, Lau *et al.* has presented a noncontact method to track the cardiac surface. But for ECET map overlay there was several acute limitations; endoscope image must contain coronary artery tree, surgical instruments can not occlude heart and specular reflections can cause instabilities. We overcome these technical limitations using the Conditional Density Propagation algorithm developed in Sec. II 2.1. On the other hand, Ginhoux *et al.* present a predictive visual servoing system trained with data from optical markers. Attaching markers is not ideal, but their approach provides greater stability and occlusion resistance. Their adaptive observer requires the use of a 500Hz video camera. This is an expensive and non-standard tool in the operating theatre. We have overcome this engineering challenge by simplifying the tracking algorithm, Sec. II 2.2, and using a fast rigid-registration method, Sec. II 2.4.

2 Method

2.1 The Conditional Density Propagation Algorithm

The Condensation algorithm provides a method to estimate a generic state, x_t and is made up of three distinct steps.

Selection Step. In the application described in this paper, the openCV Condensation estimator was used. OpenCV internally controls this selection step based on several initialization parameters. We performed some preliminary testing and found random sampling to within ± 25 pixels provided suitable tracking for our system. Smaller values provided faster processing speed but quick movements then resulted in instability.

Prediction Step. The Prediction step involves the use of a dynamic model to estimate state, x_t. In this publication state will refer to estimated marker position in frame t. Therefore, x_t is a 2D vector, $[u, v]^T$, in image coordinates. The condensation algorithm's probability frameworks make assumptions such that the dynamic model can be represented numerically as,

$$x_t = A \cdot x_{t-1} + \omega_t \tag{1}$$

where ω is some random noise function. For the Cardiac Electrophysiologic Mapping Overlay System, we let A equal the unity matrix. This means, our predictive model assumes marker position, x_t, does not change from the position in the previous frame, x_{t-1}. While it appears this would deadlock the system, the noise function, ω, accounts for marker movement and enables tracking. This simple first order model is acceptable as it was for similar work performed by Nummario *et al.* [8].

Measurement Step. Condensation implements factored sampling to perform the measurement step. For this step, a probability density function, $p(z_t|x_t)$, needs to be specified. Normally $p(z_t|x_t)$ requires either posterior density data, $p(x_t|z_t)$, to be defined or, time-consuming iterative calculations in the case of two-dimensional images [9]. However, we have found a way to avoid this in the case of the ECET Map Overlay System by reducing the problem to a one-dimensional tracker.

2.2 Probability Density Function

In the case of a one-dimensional Condensation state estimator, the observations reduce to a set of scalars [9]. This idea provides an economical alternative to the two-dimensional case. In the application described in this paper, the probability density function is transformed first into RGB color space. This is achieved by realizing a specific observation from the observation set, $z_t^{(m)}, m = 1, 2, \ldots M$ is equal to a vector in RGB color space. Succinctly, we can use $z_t^{(m)}$, a point in image space, to extract pixel information $c_t^{(m)}$, a point in RGB color space, and observe the probability this pixel is a target color, c_{target}. The observation should result in a scalar, c_m. For this we use euclidean distance between the target color and the sampled pixel. Combining these steps we can derive the following probability density function.

$$p(z_t|c_t) \propto 1 + \frac{1}{\sqrt{2\pi}\sigma\alpha} \sum_M \exp\left(-\frac{{c_m}^2}{2\sigma^2}\right) \tag{2}$$

In (2), the parameters σ and α must be chosen after examination of data and measurement error. σ is the standard deviation of the data set and we found it to be around 50 during laboratory calibration. α is less well defined but we found good results setting it to equal one. Notice we are now only tracking color space and positional accuracy is lost. We have overcome this problem using something we have called *Center of Gravity Compensation*

2.3 Center of Gravity Compensation

The condensation algorithm described in Sec. 2.1 & Sec. 2.2 allows us to quickly track the optical markers in a windowed endoscope stream. But the method has a weakness. The method is tracking the optical markers as a point in RGB color space but this transforms to a *region* in image space. The algorithm does not allow us to specifically track the center of this region, which is where we will assume the marker is located. To locate the center of the marker, we are using the C++ image segmentation library released by Imura Masataka [10]. We use the result of the condensation algorithm to isolate the marker region in the labeled image. Then we calcluate the center of gravity of that region and set this as the real marker position.

2.4 ECET Map Overlay

Augumented Reality (AR) can be used to provide information to the surgeon in an intuitive way [11]. We chose a basic multimodal AR overlay method after balancing speed and simplicity with performance requirements. We overlay epicardiac electrophysiologic data onto the endoscope frame to create a registered image which is presented to the surgeon via a monitor. The process is a rigid registration which is described below.

Feature Detection. The key to applying Equation (2) is selecting a color that is distinguishable from image clutter. *In vivo*, the system is targeted to function using surgical images delivered via an endoscope. Obviously, colours such as black, red and white are abundant in such images. For this reason marker color, c_{target}, was chosen to be blue.

Feature Matching. Feature matching is performed using a Mutual Information (MI) method as is common with multimodal registrations [12]. We use *a priori* knowledge of the relationship between the electrode array and markers to calculate the correct position for overlay. We have documented an electrode array and ECET Map generation in a previous publication [13]. However, in this new system we are using a global electrode array which has a fixed spatial relationship between each electrode. We also know the overall height & width. To register this map with cardiac movement, we placed a marker on the epicardium at each corner of the electrode array. Each corner marker was then tracked using the Condensation algorithm. We then use the pose of the four markers as points to 'attach' the corner of the ECET Map to in the endoscope stream. This is a form of rigid registration based on a first order global polynomial mapping function.

2.5 System Configuration

System configuration is given in Fig. 1. First the processing PC must collect electrophysiologic information and process this offline to create an ECET map. This is performed by using the Polaris Optical Tracking system by NDI to track optical markers attached to the distal end of a modified 1DOF forcep. Onto the proximal end of the forcep we attached a custom designed global electrode array to record electrophysiologic information. ECET map overlay only involves estimating the current state, registration and then overlay. Registration is easy because the four blue-markers were located using a Polaris pen tool (8700340) [1]. This allows us to accurately know the relationship between the electrode array and the four markers. Then, electrophysiologic data segmentation is performed by thresholding the measured potential with some value, T. If the recorded potential is above T then we assume it is part of the propagation wave. The

[1] In detail, we assumed the position of each marker to be the average of pen tip data after touching the center of each marker for 10 seconds.

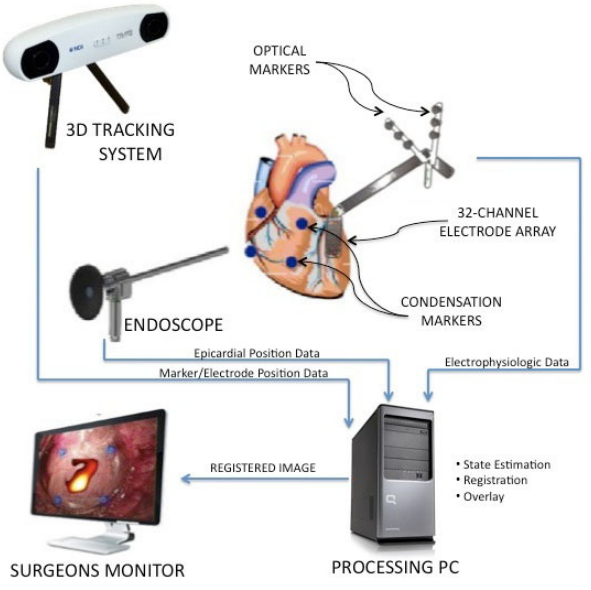


Fig. 1. System configuration. Blue markers are tracked using endoscope stream while electrode position is recorded using a 3D tracking system. The two are registered and then displayed on a monitor.

result is a mask bitmap stream which can be used to decide if output pixel, $G(i,j)_t$, will come from the endoscope stream, $F(i,j)_t$, or be bit copied from the ECET map, $D(i,j)_t$. Eq. 3 defines the process mathematically,

$$G(i,j)_t = \begin{cases} F(i,j)_t \text{ if } D(i,j)_t < \mathrm{T} \\ D(i,j)_t \text{ if } D(i,j)_t \geq \mathrm{T} \end{cases} \tag{3}$$

3 Laboratory Evaluation

3.1 Tracking Accuracy

Evaluation of algorithm accuracy was performed in image space. Two experiments were performed to evaluate two parameters. The first experiment was to determine tracking accuracy as the target moves increasingly faster, the second experiment was to determine tracking accuracy as the tracking markers became smaller. Both experiments used a simulation program to feed the condensation algorithm with software controlled markers.

Tracking Accuracy vs Target Speed. To obtain quantitative results of algorithm accuracy, four markers were moved in a circular trajectory. The marker speed was made controllable via keyboard input. In this way, we could know the

exact position of each marker. The next step was to pass this bitmap into the condensation algorithm and calculate the estimated position. Now, we could also know the estimated position of each marker. The results of one such experiment are given in Fig. 2.

Tracking Accuracy vs Marker Size. *In vivo* it is desirable that the tracking algorithm can track the epicardium via endoscopic images. Consequently, marker size can be deminished through blood contamination, obstruction via surgical instruments or simply by zooming out. To evaluate robustness, accuracy was

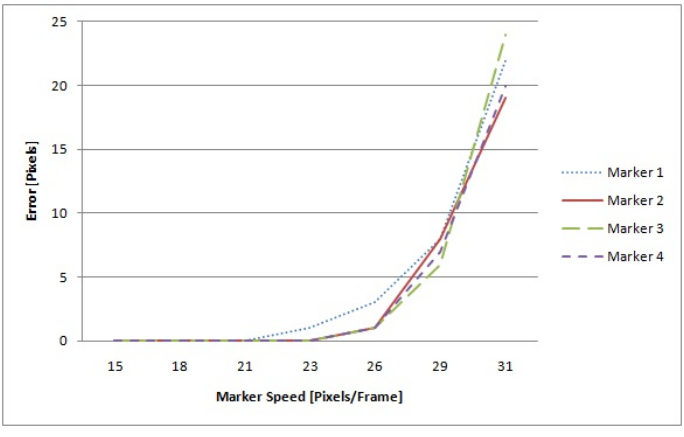


Fig. 2. Example graph showing increasing tracking errors as each marker speed increases. Performed with marker radius fixed at 5pixels.

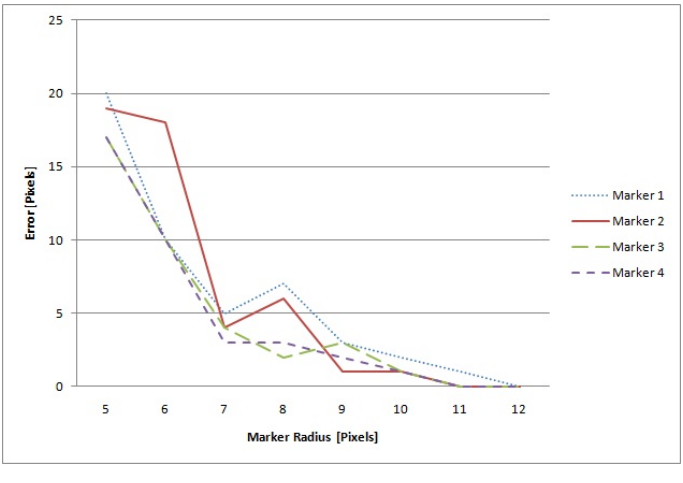


Fig. 3. Example graph showing decreasing accuracy as marker size decreases. Performed with marker speed fixed at 31pixels/frame.

measured as marker size was reduced. The simulation software described earlier was again used, however this time the keyboard input controlled marker radius. The results of one such experiment are given in Fig. 3.

4 In Vivo Evaluation

4.1 Experiment Environment

The epicardium tracking system was tested *in vivo*. The subject was a supine 39.5Kg swine, and underwent a median sternotomy with thoracotomy incision. Electrophysiologic reference potential was measured from a clip secured on the inner side of the incision site. Both this reference potential and the 20-channel epicardial electrogram, captured from our custom electrode array, were input into a multi-channel electrogram recording system (EP-WorkMate, EP Medsystems). Once collected, the data was converted to an ASCII character stream and processed in Matalab R2009a to create our ECET maps. To track heart movement, four blue markers were attached to a silicon sheet, Fig. 4. This sheet was then sutured onto the epicardium around the area to be mapped. A standard 640x480 mono-channel endoscope stream was captured using a PCI5531 image capture board by Interface. The processing PC was an Intel Core 2 Duo 2.66GHz Processor, running a MSVC2006 Debug build under Windows XP SP2.

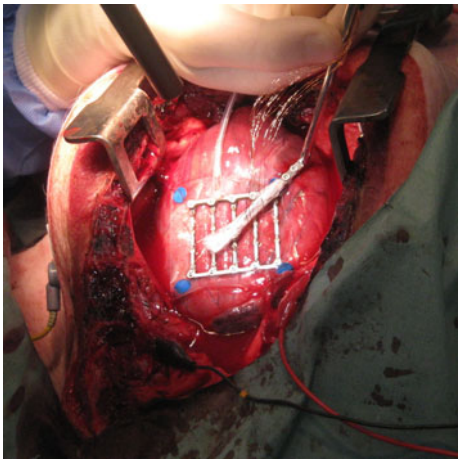


Fig. 4. Experiment environment showing global electrode array and condensation markers sutured to the epicardium.

4.2 Experiment Result

30 seconds worth of electrophysiologic maps were recorded and then processed offline. When processing was complete the maps were overlaid onto the endoscopic

image using the method described in Section 2.4. One cycle is given (every second frame) in Fig. 5. It is possible to see how effective the estimator was at picking up the markers in a dynamic, reflective and cluttered environment. Tracking took 32.73ms per frame (average over 30 second recording). This equates to a 30.6Hz frame rate. This is fast enough to operate at NTSC video speed and we consider this result to be real-time capable. Removing debug code and switching operating system provided a speed enhancement again.

Fig. 5 shows a regular (non-arrhymogenic) wave propagation across the right ventricle. The wave propagates from the bottom right of the image to the top left of the image. This makes sense because anatomically ventricle propagation originates in the apex of the myocardium and moves upwards through the ventricular myocardium [1]. This corresponds to the contraction of the ventricles pushing blood through the semilunar values; the QRS wave in an ECG recording. We used this quanlitative analysis to conclude correct wave propagation is being displayed.

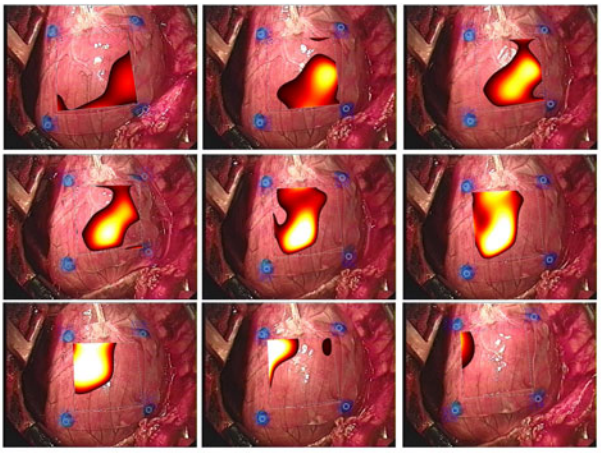


Fig. 5. Cardiac Electrophysiologic Mapping Overlay system showing wave propagation across right ventricle

5 Discussion

Condensation is one example of a Particle Filter. Particle Filters are a family of model estimation techniques based around simulation of Bayesian models. There are several alternative estimators that solve a comparable problem. Several are noted here.

Kalman Filter. Perhaps the most generic approach is to implement a Kalman filter. Unfortunately, Kalman filtering suffers several chronic impediments in the case of the Cardiac Electrophysiologic Mapping Overlay System. Firstly,

the Kalman filter is based on Gaussian densities. Gaussian densities are unimodal and cannot represent simultaneous alternative hypotheses. So if background clutter density is temporarily greater than marker density, the system will start tracking the clutter. Particle filters, can track multiple markers simultaneously, therefore reducing the chance of losing the marker. Furthermore, the Kalman filter involves solving a Riccarti Equation of considerable computational complexity. In a real-time application this is of notable concern.

Monte Carlo Methods. Monte Carlo Methods [14] [15] also use Factored Sampling in the Measurement Step. As expected, the result is a state estimator that has similar performance metrics as the Condensation algorithm. The only noticeable difference is speed of operation. In theory the Condensation algorithm can achieve comparable accuracy but faster.

6 Conclusion

We have developed a method to display epicardial excitation time maps intuitively to the surgeon. This is a new way for thoracic surgeons to locate arrhythmogenic tissue or abnormal conduction pathways to be targeted with catheter ablation. The clinical application of such a system is minimally invasive surgery compatible inter-operative and post-operative evaluation. Another contribution of this research is a method of stable and fast cardiac cycle tracking via the surgeons endoscopic stream. This is applicable in many guidance based systems which have or are currently being developed.

Acknowledgements

Part of this study was funded by NEDO P10003 Intelligent Surgical Instruments Project, Japan.

References

1. Tortora, J.G.: Principles of Human Anatomy, 9th edn. John Wiley & Sons, Inc., Chichester (2002), ISBN 0-471-38728-2
2. Benjamin, E., Wolf, P., D'Agostino, R., Silbershatz, H.: Impact of atrial fibrillation on the risk of death: The framingham heart study. Journal of The American Heart Association (1998)
3. Wattigney, W., Mensah, G., Croft, J.: Increased atrial fibrillation mortality: United states, 1980-1998. Americal Journal of Epidemiology (2002)
4. Cox, J., Schuessler, R., D'Agostino, H., Stone, C.: The surgical treatment of atrial fibrillation. iii. development of a definitive surgical procedure. J. Thorac. Cardiovasc. Surg. 101, 569–583 (1991)
5. Ginhoux, R., Gangloff, J.A., de Mathelin, M.F.: Beating heart tracking in robotic surgery using 500hz visual servoing, model predictive control and an adaptive observer. In: IEEE Conf. on RAS (April 2004)

6. Szpala, S., Guiraudon, G., Peters, T.: Real-time fusion of endoscopic views with dynamic 3d cardiac images: A phantom study. IEEE Tran. Medical Imaging (2005)
7. Lau, W., Ramey, N., Corso, J., Thakor, N.: Stereo-based endoscopic tracking of cardiac surface deformation. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3217, pp. 494–501. Springer, Heidelberg (2004)
8. Nummiaro, K., Koller-Meier, E., Van Gool, L.: An adaptive color-based particle filter. Image and Vision Computing (2002)
9. Isard, M., Blake, A.: Condensation - conditional density propagation for visual tracking. International Journal of Computer Vision 29, 5–28 (1998)
10. Masataka, I.: Labeling.h (July 2005), http://oshiro.bpe.es.osaka-u.ac.jp/people/staff/imura/index_html
11. Azuma, R.T.: A survey of augmented reality. Teleopertators and Virtual Environments 6, 355–385 (1997)
12. Zitova, B., Flusser, J.: Image registration methods: a survey. Image and Vision Computing (2003)
13. Yuhei, T.: Global epicardial electrophysiological mapping with local multi-channel electrode array. JSCAS 11(1), 25–38 (2009) (in Japanese)
14. Kitagawa, G.: Monte carlo filter and smoother for non-gaussian nonlinear state space models. Journal of Computer Graphic and Statistics 5, 1–25 (1996)
15. Bolviken, E., Acklam, P., Christophersen, N.: Novel approach to non-linear/non-gaussian bayesian state estimation. IEEE Proceedings for Radar and Signal Processing 140, 107–113 (1993)

Knowledge-Based Situation Interpretation for Context-Aware Augmented Reality in Dental Implant Surgery

D. Katić[1], G. Sudra[1], S. Speidel[1], G. Castrillon-Oberndorfer[2], G. Eggers[2], and R. Dillmann[1]

[1] Institute for Anthropomatics (IFA), Humanoids and Intelligence Systems Laboratories (HIS), Karlsruhe Institute of Technology (KIT), Germany
darko.katic@kit.edu
[2] Department of Cranio-Maxillofacial Surgery University of Heidelberg, Germany

Abstract. The objective of this research is to develop and evaluate a context-aware Augmented Reality system which filters content based on the local context of the surgical instrument. We optically track positions of the patient and the instrument and interpret this data to recognize the phase of the operation. Depending on the result, an appropriate visualization is generated and displayed. For the interpretation, we combine a rule-based, deductive approach and a case-based, inductive one. Both rely on a description-logic based ontology. In phantom experiments the system was used to support implant positioning in models of the mandible. It recognized the phase correctly and provided an appropriate visualization about 85% of the time. The knowledge-based concept for intraoperative assistance proved capable of generating useful visualizations in a timely manner. However, further work is necessary to improve accuracy and reduce the deviation from the actual and planned implant positions.

1 Introduction

Intraoperative Augmented Reality (AR) has shown great potential in simplifying the performance of surgeries [1]. However, most existing systems only offer assistance for a single, predefined phase of the operation. Incapable of adjusting to current events, the visualizations can even cause problems by obstructing the surgeon's view during sudden emergency situations. Therefore, there is research interest in analyzing the workflow to offer assistance depending on the current context [2,3].

In dental implant surgery, there are times when the surgeon just needs to see the position of a single dental implant positions. And sometimes he additionally needs to see the positions of nearby vital structures. The idea is that at any point in time, just the information currently relevant to the surgeon should be displayed, while everything else is filtered out. We propose a knowledge-based method to provide this service.

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 531–540, 2010.

For context-aware visualization, situational awareness is required. Commonly, stochastic classifiers such as HMMs or neuronal networks [4,5,7] are used. Other systems use knowledge representation techniques to model and analyze the surgical workflow [6]. In [8] we presented an intraoperative knowledge-based approach for context-aware systems based on the work of Neumann et al. [9]. Its use of ontologies allows for formally sound and reusable incorporation of background-knowledge. All information is stored in a human-readable white-box fashion. It can easily be checked for soundness, either manually through experts or automatically through formal verification algorithms. In medical applications with a focus on safety, this is a major advantage compared to stochastic methods. Their numerical representation of knowledge is rarely accessible to human interpretation and understanding.

The basic idea behind our situation interpretation, as introduced in [8], is to use the same information sources a surgeon also would: formal knowledge, as found in literature, and the experience he acquired in his working life. For dental implant surgery, we improved our methods and developed an AR-based planning tool for dental implants. The algorithms are implemented in a system for intraoperative assistance, which has already been used in a clinical study [10]. The methods and the results of phantom experiments are presented in this paper.

2 Methods

2.1 Preoperative Planning

Precise planning of implant positions is of great importance for dental implant surgery. For our system, implant positions can be defined with common planning tools. However, there are few systems that go beyond traditional WIMP-interfaces (Windows, Icons, Menus and Pointing). To evaluate the benefits of alternative interfaces, we developed a planning tool based on AR.

The planning tool overlays DICOM-image slices on a 3D polygonal model of the mandible. User interaction takes place via a 3D pointing device with integrated buttons. They are used to scroll through the image slices and define implant positions. Examples are shown in Fig. 1.

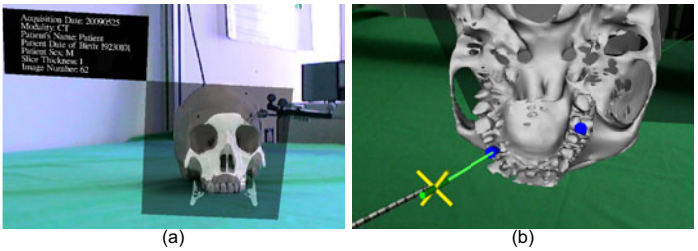


Fig. 1. (a) skull model with a CT-image slice, (b) positioning of implant positions represented with blue cylinders

Our approach allows for intuitive navigation and planning. The user can physically place and rotate objects. This is a major benefit in comparison to the often cumbersome WIMP-based methods for defining positions of implants. The tool is integrated in the same software framework which also hosts the modules for intraoperative assistance. This way, migration from planning data to its intraoperative use is simplified. Further work will focus on improvements of the user interface to ensure a more streamlined workflow.

2.2 Intraoperative Assistance

Intraoperative assistance is provided in a three-step process: situation modeling, interpretation and visualization. Each of them is described in the following.

Situation Modeling. The aim of this step is to create a computational model of the current situation to serve as the basis for the interpretation. It is compromised of preoperative planning data and intraoperatively acquired information on positions and distances. To account for differing requirements, three different models of the situation are used: the subsymbolic, symbolic and semantic model.

The subsymbolic model represents the current situation in a quantitative way. Objects are represented as polygonal meshes and relations between them, e.g. *distance*, as numerical values. This is the most detailed model of the situation. Its purpose is to provide parameters for the visualization and it forms the basis for the symbolic model. In the current system, we use the NDI Polaris optical tracking device as the only sensor. It allows for tracking of features like position, speed and distance in real-time.

The symbolic model represents the current situation in a qualitative way. Objects are represented with symbolic names, e.g. *drill* or *nervusAlveolarisInferior*, relations with symbolic qualifiers like *near* or *far* for the distance relation. This text-based representation offers a concise view on the current situation. Currently, we track two low-level features: a measure of distance and a measure of approaching. Approaching is quantified by a combination of approaching speed and distance. Object which are already in proximity of each other and whose distance is decreasing quickly are assigened a high degree of approaching. The corresponding function was defined using fuzzy logics. It is defined as the conjunction of the fuzzy predicates *proximity* and *decreasing distance*. The T-Norm min is used as the conjunction operator.

The semantic model represents the current situation with Description Logics (DLs). Here, knowledge about the current situation is augmented with medical background information. As common in DL, knowledge about the current situation is expressed in the Assertional Box (ABox), whereas background knowledge is modeled in the Terminological Box (TBox). For the problem at hand, we use a new means of representation in comparison to [8]. The situation is described by binary relations between the drilling device and all other objects (namely the drill burs, anatomical structures and implant positions). An example is provided in Fig. 2. The semantic model is used as the basis for knowledge-driven situation interpretation. The logic reasoning is performed by RacerPro, a commercial reasoning system for OWL, on a handcrafted Ontology.

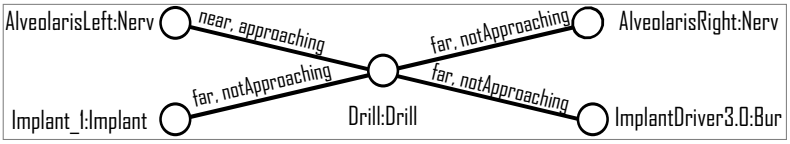


Fig. 2. Small excerpt of an ABox, showing the relations between the drilling device and four other objects

Situation Interpretation. The purpose of the situation interpretation process is to determine the current phase of the operation. We mimic the human approach to this problem by using formal knowledge from literature and experience-based knowledge from observed examples. To represent formal knowledge, we use a rule-based system. Experience is modeled by case based reasoning. Using two complementing methods, we increase confidence in the classification results. Both algorithms rely on the semantic model as a common knowledge base and take advantage of logical inference mechanisms. Currently four phases of the operation are considered: drilling, picking a new bur, approaching a planned implant position and endangering the nervus alveolaris.

As introduced in [8], the rule-based approach relies on explicit definitions of phases with rules of canonical "if-then"-structure. For technical reasons, we use nRQL (new Racer Query Language) to express rules in terms of queries. The queries consist of the conditional part of the rule, the answers to queries of all objects which satisfy these constraints. If the answer set is empty, we know that the corresponding phase is not valid, otherwise it is. For instance the rule "if a sharp instrument is near a vital structure, then there is a risk situation" can be expressed as the query "retrieve all vital structures and sharp instruments that are near each other". This way the current situation can be classified.

Building on the idea of Case Retrieval Nets [13], we define situations implicitly with a set of labeled training samples. These samples are called cases. A case consists of a description of the situation and a label indicating its phase. The basic units of situation descriptions are Information Entities (IEs). They specify the relation between two objects and are used to express observations. For instance the observation of a drill in proximity of a planned implant position is described by the IE *(Drill, near, Implantposition)*. Several cases are collected in a case base and used to classify new situations (so called targets). New situations are assigned to the phase of the most similar case from the case base. In the following, a method to build a case base for situation interpretation is presented.

Training samples for the case base are acquired intraoperatively by a human operator or generated synthetically according to probability estimates by experts. For the interpretation process, a soft measure is computed, which specifies how much a target and a case have in common. In [11] we examined purely knowledge-driven object-to-object similarity measures. For the problem at hand, we have investigated a different method based on machine learning. To incorporate contextual knowledge we use a relevance measure derived from training

samples. It signifies the importance of an IE in regard to a case. For instance, the IE *(drill, near, ImplantDriver4.0)* is of low relevance during a drilling phase. But it is of great importance whenever a particular drilling bur is picked. Basically, the idea is that if the IE occurs often in cases for a certain phase and seldom in others, or vice versa, it is of high relevance for that phase. To realize these ideas we combined methods from information theory and Bayesian estimation.

Formally, we use the training samples to learn a function $r(i, p)$ that assigns the relevance of the IE i to a phase p. The function is defined as:

$$r(i,p) = (H(p) - H(p|i))(P(p|i) - P(p)). \tag{1}$$

The first term is known as the mutual information, with H defined as the Shannon entropy. It is a measure of mutual dependence of the phase and the IE. The higher the value, the greater the connection between them. It is undirected, giving no information about whether the occurrence is a cue for or against the phase. This information is added by the second term. If the a priori probability is greater than the probability after the IE was observed, the occurrence of this entity hints that the situation does not belong to this phase. In this case, the function returns negative values. The same lines of thought apply for the other cases. In literature, this function is known as cue validity.

As is expected and evidenced in empirical trials, there are numerous IEs that have very small yet non-zero relevance. These low relevances originate from noise in the training data. Using these IEs in the interpretation process would only increase run-time and add nothing to the accuracy of interpretation results. In fact, it might even have negative impact on the recognition rate, since they only add noise and no additional information. Therefore IEs with low relevance are filtered out, leading to a more dense and reliable case base.

Given $r(i, p)$, it is possible to compute the degree of commonality between a case c and a target t. This value is called the case-activation $a_{case}(t, c)$. With $p(c)$ defined as the phase, c belongs to. It is computed as follows:

$$a_{case}(t,c) = \sum_{i \in t \cup c} \sigma(i,t,c) r(i,p(c)). \tag{2}$$

The function $\sigma(i, t, c)$ is given by:

$$\sigma(i,t,c) = \begin{cases} 1 & (r(i,p(c)) > 0 \wedge i \in t \wedge i \in c) \vee (r(i,p(c)) < 0 \wedge i \in t) \\ -1 & r(i,p(c)) > 0 \wedge i \notin t \wedge i \in c \end{cases} \tag{3}$$

The difference between the case and target is calculated and weighted with the corresponding relevance. If an IE indicates that case and target are similar, the activation is increased. Otherwise it is decreased. The amount of change is determined by the absolute value of the relevance. If case and target only differ in irrelevant IEs, the activation is high, otherwise low.

To classify new situations, the phase-activation $a_{phase}(t, p)$ of a target t to a phase p is determined. It is defined as follows:

$$a_{phase}(t,p) = \max_{c \in p} \{a_{case}(t,c)\}. \tag{4}$$

The idea, motivated by the work introduced in [12], is to have activation values which are as high as possible but not higher than is justified by the given observations. The max-function satisfies this requirement. If a function is chosen that yields to higher values, further background knowledge is necessary to justify the phase activation since it is greater than all individual case activations. Functions with smaller values are overly conservative because higher case activations were observed. Finally, the target is assigned to the phase with the greatest activation.

Situation Visualization. In [8] we statically assigned a predefined visualization to each phase of the operation. While feasible, this method requires all phases to be known in advance. Proper visualizations need to be hand-crafted in foresight. Dealing with situations which overlap in time or only differ in small details leads to bloated and redundant visualization sets.

To deal with these problems, we developed an automated visualization construction kit. Depending on the current phase, visualization primitives are automatically assigned to each object. The visualization for the whole scene is thus defined implicitly. The assignment is made using logical inference. Visualizations are defined in the TBox as sub-concepts of *visualization*. During run-time, the most specific concept of each *visualization*-instance connected to an object is determined and the corresponding visualization is applied. Similarly to the situation interpretation, the knowledge-based approach allows for reusable and safe incorporation of background knowledge. The process is shown in (Fig. 3).

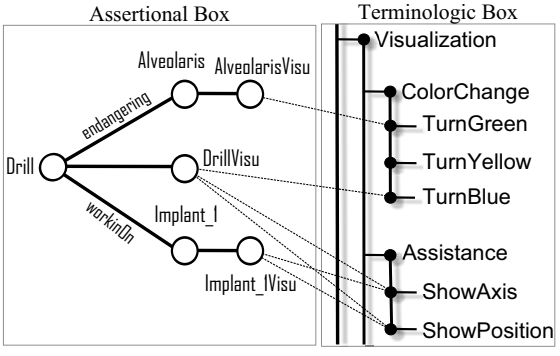


Fig. 3. By computing the most specific concepts of the *visualization*-instances, visualizations for all objects are determined

For the application in dental implant surgery, we designed the visualizations with regard to psychological theories of visual attention. When a hole is being drilled the axis of the planned implant and the dental bur is displayed, along with information about the desired and reached penetration depth. After a first phantom-experiment we extended the visualization with a view-dependent top-down display of the scene to make orienting the drill easier. When the nervus

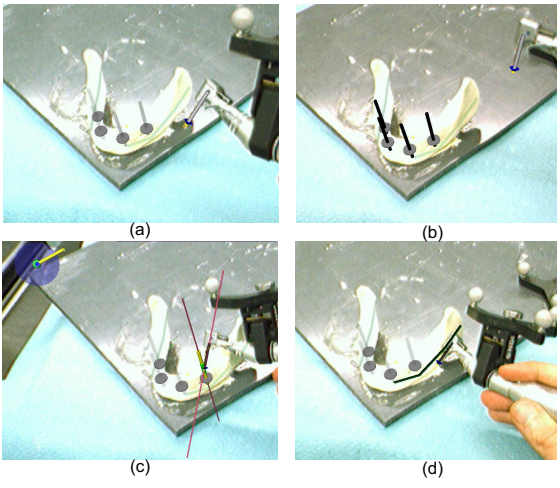


Fig. 4. The figure shows the initial visualization during idle phases (a), the approaching of planned implant positions (b), the drilling (c) and a risk situation (d)

alveolaris is endangered by a sharp instrument, it lights up in a green color. To indicate the severity of the risk situation, the color changes with the distance of the sharp instrument from the vital structure. Examples are shown in (Fig. 4).

Combing the methods in an agent-based framework. The algorithms are implemented within agents, organized in a common environment and able to communicate among each other. Two agents are in charge of determining the current phase of the operation, one employing the case based approach, and the other the query based one. A third agent integrates the answers of those agents into a single classification. It also logs all detected situations with a timestamp to create a HTML-based protocol of the surgery. A fourth agent determines an appropriate visualization.

3 Results

To evaluate recognition rate, medical usability and accuracy of the system, we conducted two phantom experiments as well as several questionnaire-based user evaluations. The experiments consisted of drilling planned implant positions in several artificial models of the mandible. The system we used in the second experiment had several improvements over the first version. The most obvious change was the use of new AR-goggles. For the first experiment we used a Sony Glasstron head mounted display (HMD). Its major flaws were low contrast displays and the severe darkening of the view on the real world. For the second experiment we used a novel device made by Trivisio. It suffered considerably less from these problems and had the advantage of a frameless design. This allows for a clear, unobstructed view on the patient, thus improving safety. Both

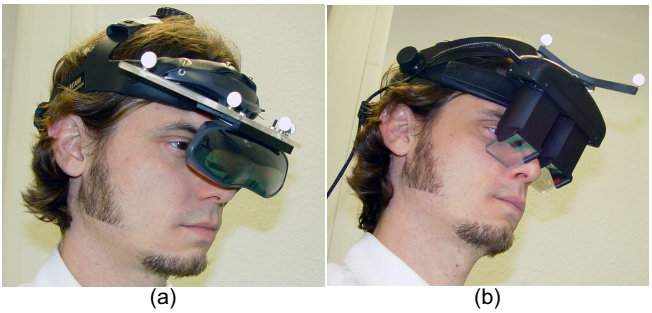


Fig. 5. The Sony Glasstron (a) and the new, frameless Trivisio design (b)

goggles are shown in Fig. 5. The results of the experiments are presented in the following.

Situation Recognition Rate. A recognition rate of over 85% was achieved. It has to be noted that real-time constraints were considered. Correct but late visualizations were considered as false recognitions. The majority of false recognitions occured during the approaching of the implant positions. The visualizations sometimes disappeared during the approach because the drill's speed dropped below a certain treshold. The system then assumed that no approaching takes place. Similarly, the minimal distance between the drill and the implant position necessary to activate the visualization for drilling was too short. Apart from maladjusted parameters, run-time characterstics were a source of broken real-time constraints and thus false recognitions. For instance, the visualization of implant postition stayed active for a while although the drill was removed from the area. This was due to the latency of the situation recogntion.

Medical Usablility. To evaluate the medical usability, the surgeons were asked to answer a questionnaire after each experiment. The feedback we received was favorable. For the most part, the system fit well in the existing workflow and provided quick and reliable assistance. Nonetheless, two flaws were found. The surgeon wanted see the axis of the planned implant and the current axis of the drill from a different angle. An additional top-down view was later added and used in the second experiment. It made orienting the drill easier. Secondly, the method to choose a new bur by touching it with the drill showed to be impractical. At this point the workflow of the surgeon was interrupted. This shortcoming needs to be addressed in future works. In summary, the surgeons confirmed that the system facilitates the performance of the surgery, yet some practical problems still remain to be solved.

Accuracy. After the experiments, the actual implant positions (three in the first experiment, four in the second per phantom) were compared to the planned ones. The average deviations are shown in Tab. 1 and Tab. 2. The results are not

satisfactory for the problem at hand. Reasons for the deviations include errors in the registration of the patient and the instrument, calibration of the HMD and measurement noise. The surgeon in the second experiment had generally better results for the orientation of the implant. This is due to the additional top-down view on the scenery which made orienting the drill easier. Yet further work is still necessary to improve accuracy.

Table 1. Deviations in the first trial

phantom	position	depth	orientation
1	3.59 mm	2.73 mm	6.54°
2	1.61 mm	0.27 mm	5.56°
3	2.26 mm	0.58 mm	3.35°
4	0.82 mm	0.54 mm	1.68°

Table 2. Deviations in the second trial

phantom	position	depth	orientation
1	2.78 mm	3.52 mm	4.36°
2	3.28 mm	3.10 mm	1.52°
3	2.11 mm	2.91 mm	4.28°

4 Conclusion

We presented a knowledge-based approach for a context-aware intraoperative assistance system. Without user interaction, our system constantly offers the currently needed assistance. Appropriate AR visualizations are provided during the entire operation, which is particularly beneficial during unforeseen emergency situations.

The situational awareness is implemented using an ontology. We merge heterogeneous data compromised of medical background knowledge, preoperative planning data and interoperativ sensor information into a common representation. Four agents collaboratively interpret the collected data, recognize the situation and provide an appropriate visualization.

The system was evaluated in two phantom experiments. Recognition rate, accuracy and medical usability were examined. In summary, the system proved capable to reliably recognize situations and provide helpful visualizations in real-time. Problems with an unsatisfactory accuracy however still remain open. Future work will focus on improving the recognition rate and accuracy as well as evaluating the concepts on more complex surgeries.

Acknowledgement. The presented research is supported by the German Research Foundation (Research Grant DI 330/23-1).

References

1. Sielhorst, T., Feuerstein, M., Navab, N.: Advanced Medical Displays: A Literature Review of Augmented Reality IEEE/OSA Journal of Display Technology. Special Issue on Medical Displays (2008)
2. Navab, N., Traub, J., Sielhorst, T., Feuerstein, M., Bichlmeier, C.: Action- and Workflow-Driven Augmented Reality for Computer-Aided Medical Procedures. In: IEEE Computer Graphics and Applications (2007)

3. Blum, T., Padoy, N., Feuner, H., Navab, N.: Modeling and Online Recognition of Surgical Phases using Hidden Markov Models. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 627–635. Springer, Heidelberg (2008)
4. Reiley, C.E., Lin, H.C., Varadarajan, B., Vagolgyi, B., Khudanpur, S., Yuh, D.D., Hager, G.D.: Automatic Recognition of Surgical Motions Using Statistical Modeling for Capturing Variability. Medicine Meets Virtual Reality (2008)
5. Bouarfa, L., Jonker, P.P., Dankelman, J.: Surgical context discovery by monitoring low-level activities in the OR. In: M2CAI (2009)
6. Neumuth, T., Mansmann, S., Scholl, M.H., Burgert, O.: Data Warehouse Technology for Surgical Workflow Analysis. In: IEEE Computer-Based Medical Systems (2008)
7. James, A., Vieira, D., Lo, B., Darzi, A., Yang, G.: Eye-Gaze Driven Surgical Workflow Segmentation. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 110–117. Springer, Heidelberg (2007)
8. Sudra, R., Katic, D., Braun, M., Speidel, S., Castrillon-Oberndorfer, G., Eggers, G., Marmulla, R., Dillmann, R.: Wissensbasierte Modellierung und Situationsinterpretation fuer eine kontextbezogene Chirurgieassistenz. In: GI-LNI (2009)
9. Neumann, B., Moeller, R.: On scene interpretation with description logics. In: Image and Vision Computing, vol. 26 (2008)
10. Sudra, G., Marmulla, R., Salb, T., Ghanai, S., Eggers, G., Giesler, B., Hassfeld, S., Muehling, J., Dillmann, R.: First clinical tests with the augmented reality system INPRES. Studies in Health Technology and Informatics (2005)
11. Sudra, G., Becker, A., Speidel, S., Braun, M., Mueller-Stich, B., Dillmann, R.: Estimating Similarity of Surgical Situations with Case-Retrieval-Nets. In: MMVR (2009)
12. Weisbrod, J.: A New Approach to Fuzzy Reasoning. Soft Computing (1998)
13. Lenz, M., Burkhard, H.: Case Retrieval Nets: Basic Ideas and Extensions. In: 20th Annual German Conference on Artificial Intelligence: Advances in Artificial Intelligence (1996)

Scorpion Shaped Endoscopic Surgical Robot for NOTES and SPS With Augmented Reality Functions

Naoki Suzuki[1], Asaki Hattori[1], Kazuo Tanoue[2], Satoshi Ieiri[2], Kozo Konishi[3], Morimasa Tomikawa[2], Hajime Kenmotsu[2], and Makoto Hashizume[2,3]

[1] Institute for High Dimensional Medical Imaging,
The Jikei University School of Medicine, Tokyo, Japan
[2] Dept of Advanced Medicine and Innovative Technology,
Kyushu University Hospital, Fukuoka, Japan
[3] Dept of Future Medicine and Innovative Medical Information,
Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan
nsuzuki@jikei.ac.jp

Abstract. In the process of developing an endoscopic surgical robot system that adapts to NOTES (Natural Orifice Translumenal Endoscopic Surgery) and SPS (Single port surgery), by making the tip a soft tubular structure and adding an augmented reality function to the system, we were able to improve the general function of the surgical robot system. First, we added a haptic sense function to avoid breaking the soft tissue and to avoid the danger of cutting it. These occur due to the small size of the touching surface between the tip of the robot arm and the soft tissue. We were able to conduct operation by feeding back to the surgeon the force applied to the soft tissue by detecting the haptic sense of the small forceps at the tip through measuring the tension variation at the base of the wire that drives the robot arm. We also mounted various numbers of augmented reality function such as grasping the exact location of the surgical robot inside the human body and information on how the robot is reaching the location of surgery. As a result, we were able to build a system that can conduct safe surgery with the system's two main characteristics - the smallness and the high degree of freedom to move.

Keywords: Endoscopic surgical robot, NOTES, SPS, Augmented reality.

1 Introduction

We have been developing a surgical robot under a new concept using microfabrication technology and tele-presence technology [1,2]. Up until now, the basis of building a surgical robot, such as in Zeus™ and da Vinci™ [3,4], was in laparoscope surgery, and their structure was based on the robot arm controlling the laparoscope and forceps. But our surgical robot that we started to develop in the year 2000 has a tip that is small enough to go inside a human body and that part can conduct operations. Our surgical robot has an eye at the tip and robot arms on each side of the eye. It is like a small robot that can operate like human hands inside

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 541–550, 2010.

minute space. As shown Fig.1, the distal part of the robot resembles scorpion with those long two arms and a small head with eyes. For this robot, we aimed to build an endoscopic surgical robot system that can adapt to both NOTES (Natural Orifice Translumenal Endoscopic Surgery) [5-8] and SPS (Single Port Surgery). NOTES is a surgery that goes inside through the esophagus via stomach and stomach wall and conducts operation in the abdominal cavity. SPS is a surgery that penetrates through the body surface and goes into the abdominal cavity to conduct operation. In developing the endoscopic surgical robot system, we found out that it was important to develop functions to solve problems that the structure of the robot itself had. The problem lay in robot's smallness and the high degree of freedom to move. These two points were also advantages of the robot but we realized that to conduct safe surgery, we needed to complement it by other technologies.

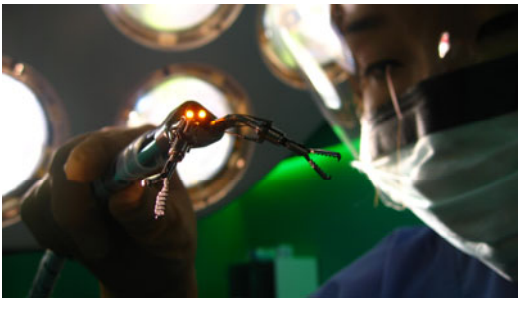


Fig. 1. Appearance of the tip of the scorpion shaped endoscopic surgical robot

First we will explain why the small size of the robot arm is a problem. As we will explain later on the method, the tip of a robot arm is 40mm in length and 6.0mm in width and is shaped like forceps. We found out that as the contact surface of the forceps and soft tissue is so small, if the force applied to the forceps is too big, it would crush or cut the mucosa layer of the stomach wall or the intestinal wall.

We needed to have the small forceps tip that would detect its haptic sense and have the amount of applied force fed back to the surgeon. In this thesis, we report on how we are in the process of developing the haptic sense function.

In addition, this surgical robot can move freely inside a human body during conducting laparoscope operation. But due to this characteristic, there were cases where the surgeon lost the location of the robot inside the body. Therefore, we found out that we highly needed to complement the robot with augmented reality technology so that the surgeon can grasp the exact location of the robot and get information on how the robot is reaching the location of surgery. Moreover, we found out that as the robot is small in size, the surgeon can easily be blinded by small amount of blood. It can also be buried in a stump of soft tissue which was not a problem in open surgery or laparoscopic surgery.

2 Method

2.1 Structure of the Robot

Before we report on the haptic sense function and the augment reality technology we used, we will explain the structure of the robot we are developing now.

Fig.2 shows the block diagram of the surgical robot system in whole. Two robot arms are mounted on the tip of the endoscopic like shape, which goes inside a human body. Each robot arm and the forceps at the tip are driven by wires from the actuator unit. The unit comprises of stepping motor assembly located at the based of the robot arms and forceps.

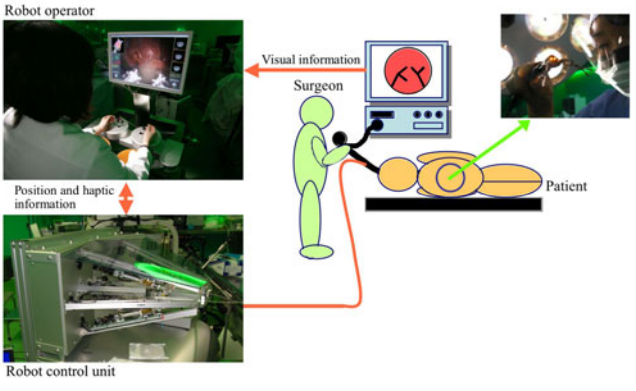


Fig. 2. Block diagram of the surgical robot system

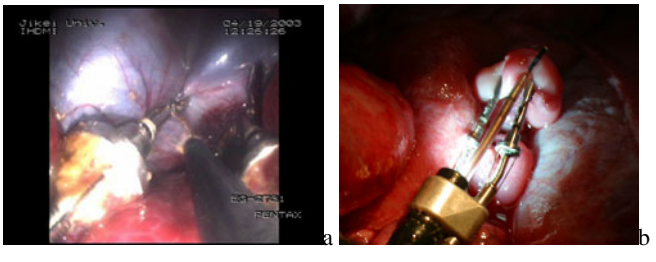


Fig. 3. Scene of the animal experiment using the endoscopic robot system. a: handling of gallbladder, b: clipping oviduct

This endoscopic surgical robot is controlled by two surgeons working in liaison. Surgeon A inserts the robot into the body and moves, rotates and changes direction of the robot near the destination site. Surgeon B sits in front of the console and controls the robot arms and conducts surgery. Surgeon A inserts and withdraws relocation clipping device and surgical needle knife using lumen connecting the tip of the robot and the outside of the body. He also cleans and conducts suctioning the lens and the location of surgery using lumen. Both Surgeons A and B can use the endoscopic

screen reinforced by augmented reality. Fig.3 shows the system during animal experiment. Fig.4 shows actuator unit that comprises of stepping motor assembly for each ten robot arms (five on each side). The unit drives the robot arms. The stepping motors in the actuator unit is positioned in a cone shape with the robot's base at the center so that that wire force will effectively be conveyed to the robot arms.

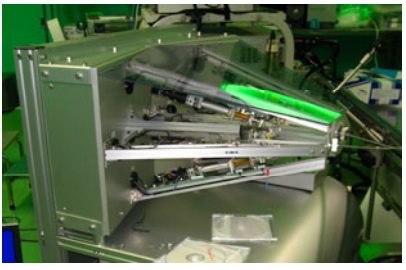


Fig. 4. Actuator unit of the robot system

2.2 Structure of Robot Arms

The robot arms on the left and right have symmetric structures. Each has four wires that move the robot arm up and down, right and left and a wire that open and closes the tip of the arm. The tip of the forceps is 10mm in length and 2.0mm in width. When opened to the limit, it can grab and object of 6mm. The tip of the robot arm can bend maximum 12mm up and down, right and left by four wires driving it in liaison. A robot arm can work in up to 40mm ground from itself, right and left. We also aimed high-efficiency for wire pulling force so that the robot arm can grab soft tissue with a force more than 3N and the forceps can open and close with that force.

Fig.5 shows the structure of the tip of the robot arm. The four wires positioned around the forceps enables the robot arm to bend up and down, right and left.

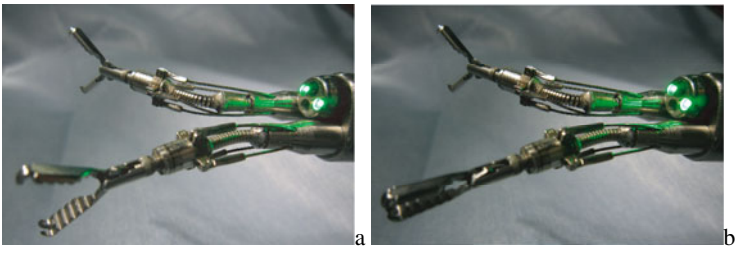


Fig. 5. Structure of the tip of the robot arm

2.3 Acquiring Haptic Sense

We at first tried to acquire haptic sense by positioning a pressure sensor at the mechanical section of forceps. But it was difficult to adequately position piezo element on the surface of the insides of minute forceps. There was also danger of

safety problems such as strength of the coat of the wire parts that would be connected to the device was not strong enough and had the danger of electrical leak.

To solve this problem, this system has a "Wire Traction Control" mechanism which monitors the traction of the wires at all times to obtain the best control system anytime.

Needless to say, the traction of the wires obtained at the base is information that includes various disturbance elements such as the change of position, loosening, and twists of the wires themselves. This will occur as to bend the endoscopic surgical robot, the wires run parallel and other wires are positioned to lead and protect them in the pipe. But we tried to distinguished the characteristics of the object the forceps grabbed from the characteristic curve of the pressure change against action. We conducted a mock experiment where the characteristic was a given phantom. We obtained the change of traction according to the change of physicality and succeeded in distinguishing the characteristics of the object. Therefore, we were able to acquire haptic sense function at the tip of the robot arm.

2.4 Integrated Display Function

We separated the monitors in parts and positioned each augmented reality information in a part. We positioned the endoscopic image at the center. Then we positioned superimposed image of the targeted part in 3D where it is difficult to see in the endoscopic image.

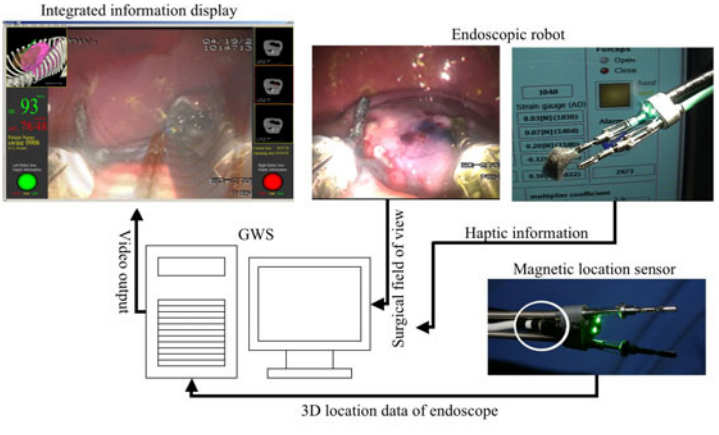


Fig. 6. Block diagram of the integrated display function

In addition, we positioned several kinds of images for navigation surgery. Moreover, we positioned display of haptic sense information of the right and left robot arms in color and the patient's vital signs. We positioned these images so that the surgeon can instinctively acknowledge and effectively use patient's information while in operation. We verified the above by conducting trial experiment with volunteers. To obtain the direction of the position of the tip of the robot arm, we mounted Minibird which is 6 DOF magnetic sensor at the lower part of the tip. We aimed to minimize the use of metal at the robot's tip and used titanium alloy so that

disturbance factor by magnetic positioning sensor metal will be small. We show in Fig.6 block diagram of the system for integrated display function.

3 Results

3.1 Acquisition of Haptic Sense

We show in Fig.7, change of traction against wire displacement when the forceps at the tip of an robot arm grabs metal, plastic eraser (soft poly-vinyl chloride), and sponge (expanded polyurethane).

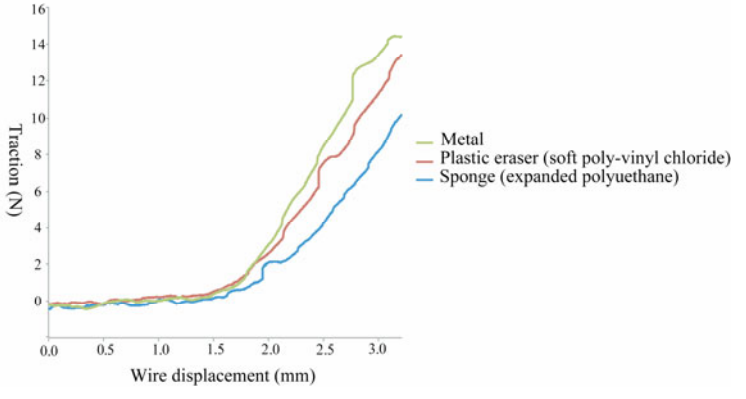


Fig. 7. Traction against wire displacement

As you can see from the change of traction against displacement of wire when the forceps is grabbing a rigid metal object, traction curve shows different characteristics from the action grabbing a rigid object in ideal condition. The following can be assumed for the reason of change displayed on the graph. When the forceps are closed, the wire is not always positioned in a line ideally. The small looseness of wire inside the pipe that protect the wires, and change of shape of the pipe itself when wires are pulled, all show up as peculiar change. The same can be assumed for an elastic body that change shape easily and either way, it looks like it is difficult to acknowledge the shape of the object the arm has grabbed. But as can be seen in the graph, the characteristics of the object the forceps have grabbed showed as differences of change of traction.

In addition, we found out that if we know the softness of the object when grabbing it, we can feedback grabbing power to avoid breaking the object. For example, we were able to conduct a mock operation by feeding back to the surgeon, the force applied to the soft tissue so that he could operation by not crushing the soft tissue. We experimented by creating a mock stomach wall with soft urethane that would break easily. We conducted an experiment using these function. We mounted haptic sense function on the surgical robot system as following.

3.2 Mounting Function of Distinguishing Softness and Feedback of Grabbing Force

Fig.8 shows projected result of the softness of each sample the forceps have grabbed. The change of colors of the indicators on the monitor, which is at the back of the tip of the robot shows the projected result. The indicator turns green when the forceps grabs a very soft object, yellow when the forceps grab a moderately soft object and red when the forceps grab a hard object.

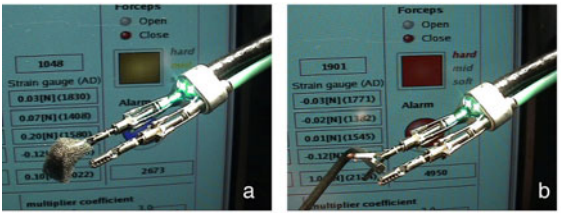


Fig. 8. Result of projection of softness as robot grabs an object. The projected result is shown in the display at the back of the robot arm. The color of the indicator changes. a: When the arm grabbed a sponge, b: When the arm grabbed a metal object.

Fig.8a shows the indicator green as the forceps are grabbing a sponge, Fig.8b shows the indicator is red as the forceps are grabbing a metal object. For functions to avoid damaging the soft tissue when the forceps grab it, Fig.9 shows the result of when the forceps grabbed silicon rubber (Fig.9a-c) which is close to the softness of soft tissues. Fig.9d-f shows the operator trying to lift up the soft tissue adjusting the forceps depending on the indicator. As mentioned before, these projected results are sent to the integrated information display system via the network and is displayed along with other operational information.

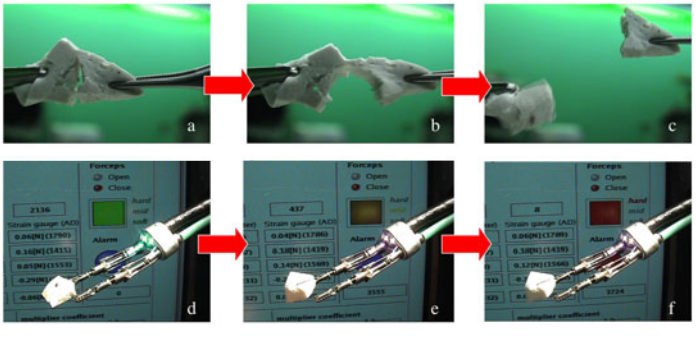


Fig. 9. Result of experiment of the robot arm grabbing the sample without damaging it. Using silicon rubber that has the same softness as soft tissue (a-c), the operator is controlling the opening and closing of the forceps by looking at the change of color or the indicator (d-f) and trying not too damage the silicon rubber by grabbing it too hard.

3.3 Integrated Display Function

As the results, the system has the following five functions. All information is updated in real-time and displayed on the screen in the cockpit for the operator. 1) function that displays the present orientation of the robot by showing patient's X-ray CT and MRI data sets superimposed onto the operation screen, 2) function that displays the position of the tip of the robot on a inner structure model reconstructed in 3D from the patient's pre-operation X-ray CT and MRI data sets, 3) function that displays the present position of the tip of the robot on patient's pre-operation X-ray CT and MRI data set images, 4) function that displays the softness of an object that the robot's manipulator has grabbed (this is being developed in parallel with the system), 5) function that displays patient's medical information such as heart rate and blood pressure.

For function 1), we use the positioning information of the robot obtained from magnetic positioning sensor installed at the tip of the robot. In this way we display in multi-layer, patient's inner structure in real-time on X-ray CT and MRI data sets obtained before operation on the operation screen. The RMS error between the endoscopic image and the superimposed X-ray CT datasets was 2.4mm.

For function 2), using the same position information at the tip of the robot as in 1), we display the position of the tip of the robot on the patient's 3D model structured before the operation. In this function, we enabled the operator to observe the 3D model from any point so that the operator can look over the whole picture to grasp the orientation of the robot and by doing so this function complements function 1).

Function 3) also uses the magnetic positioning sensor information and displays the position of the tip of the robot on X-ray CT and MRI images. The width of the window screen of the image display can be changed, the level can be changed and also the operator can enlarge and minimize the screen so that the operator needs to take his/her eyes off the operation as little as possible. Function 4) uses functions to calculate the softness of an object that a robot manipulator has grabbed. This is being developed by this project also. Function 4) can display in three different colors the softness the two manipulators each grabbed. We also created the system so that the operator can confirm patient's various medical information if it is needed during an operation in function 5).

Fig.10 shows what the system's screen displays in the operator's cockpit in an animal experiment using a pig. In the center of the screen shows function 1) mentioned earlier, at the upper left, function 2), at the upper right, function 3), at the center left, function 5) and in the lower right and left function 4) which displays the softness of the object the 2 robot manipulators grabbed. In the same figure, for function 1), the structure of the backbone and costal bones from the volume rendered X-ray CT data sets are superimposed onto the endoscope's image. For function 2), a green pointer is displayed on the surface rendered organ model and shows the position of the tip of the robot. For function 3), three consecutive images of X-ray CT images are displayed and a red cross in the center of the image shows the position of the tip of the robot. For function 4), the left is colored green and the right is colored red. This shows that the left manipulator is grabbing a soft object and the right manipulator is grabbing a hard object.

The frame rate in this experiment was 5-15 fps. For example, the frame rate was 5 fps when we used heavy X-ray CT datasets such as 512x512x512 pixels (when we used function 1).

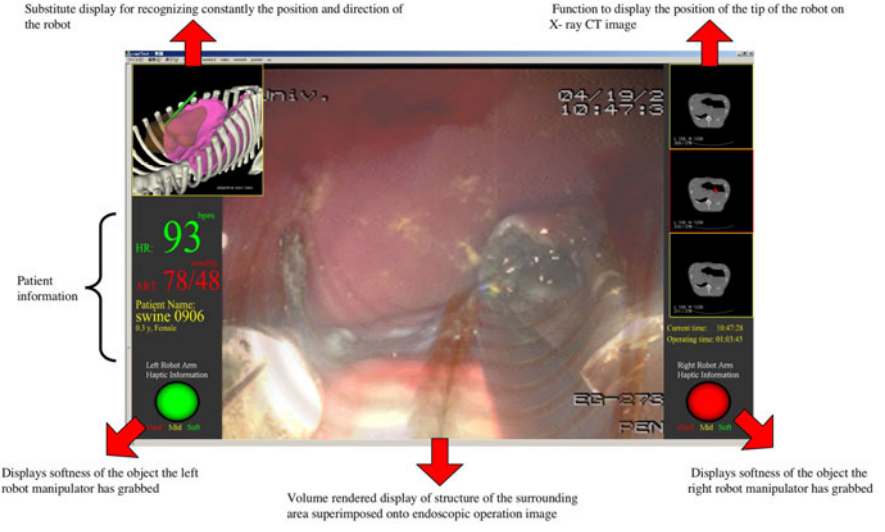


Fig. 10. Integrated information display system screen

4 Discussion

We were able to improve the function of the whole system by mounting the following function in the process of developing an endoscopic surgical robot system that adapts to NOTES and SPS. First we were able to overcome the problems due to the structure and shape of the robot system in development. We were able to reinforce the system by using augmented reality functions including haptic sense of robot arms and overcoming inadequate points such as the smallness of the robot and its freedom of movement inside the body.

As a result, we can say that by developing surgeon's console with augmented reality function including haptic sense function of the minute robot arm, we were able to build a system that can conduct safe operation with the two main characteristics of this robot system - the small size and the freedom of movement inside the body. In addition, the haptic sense device which does not need a haptic sense sensor that touches the part of operation can be applied to various other surgical robots. The integrated display system which can be display various information of the operation part including haptic sense from a various points of view can be applied to existing laparoscope type surgical robots, vascular catheter devices and radiotherapy devices. The endoscopic surgical robot system that we developed is planned to undergo clinical trials after conducting safety trials.

Acknowledgments. This study was funded by NEDO Intelligent Surgical Instruments Project.

References

1. Suzuki, N., Sumiyama, K., Hattori, A., Ikeda, K., Murakami, E.A.Y., Suzuki, S., Hayashibe, M., Otake, Y., Tajiri, H.: Development of an endoscopic robotic system with two hands for various gastric tube surgeries. Medicine Meets Virtual Reality 11, 349–353 (2003)
2. Suzuki, N., Hattori, A., Satoshi Ieiri, S., Konishi, K., Maeda, T., Fujino, Y., Ueda, Y., Tanoue, K., Hashizume, M.: Tele-Control of an endoscopic surgical robot system between Japan and Thailand for Tele-NOTES. Medicine Meets Virtual Reality 17, 374–379 (2009)
3. Reichenspurner, H., Damiano, R.J., Mack, M., Boehm, D.H., Gulbins, H., Detter, C., Meiser, B., Ellgass, R., Reichart, B.: Use of the voice-controlled and computer-assisted surgical system ZEUS for endoscopic coronary artery bypass grafting. J. Thorac. Cardiovasc. Surg. 118, 11–16 (1999)
4. Guthart, G.S., Salisbury, J.K.: The Intuitive Telesurgery System: Overview and Application. In: Proc. of the IEEE International Conference on Robotics and Automation (ICRA 2000), San Francisco CA (April 2000)
5. Piskun, G., Rajpal, S.: Transumbilical laparoscopic cholecystectomy utilizes no incisions outside the umbilicus. J. Laparoendosc. Adv. Surg. Tech. A 9, 361–364 (1999)
6. Kalloo, A.N., Singh, V.K., Jagannath, S.B., Niiyama, H., Hill, S.L., Vaughn, C.A., Magee, C.A., Kantsevoy, S.V.: Flexible transgastric peritoneoscopy: a novel approach to diagnostic and therapeutic interventions in the peritoneal cavity. Gastrointest Endosc. 61, 601–606 (2004)
7. Kantsevoy, S.V., Hu, B., Jagannath, S.B., Vaughn, C.A., Beitler, D.M., Chung, S.C.C., Cotton, P.B., Gostout, C.J., Hawes, R.H., Pasricha, P.J., Magee, C.A., Pipitone, L.J., Talamini, M.A., Kalloo, A.N.: Per-oral transgastric endoscopic splenectomy: is it possible? Surg. Endosc. 20, 522–525 (2006)
8. Marescaux, J., Dallemagne, B., Perretta, S., Wattiez, A., Mutter, D., Coumaros, D.: Surgery Without Scars: Report of Transluminal Cholecystectomy in a Human Being. Arch. Surg. 142, 823–826 (2007)

Optimisation of Focal Length Using a Stereoscopic Operating Microscope for Augmented Reality Surgical Guidance

G. González-García and R.J. Lapeer

School of Computing Sciences, University of East Anglia, UK.

Abstract. In this paper, a method is proposed that improves the focal length estimation in an augmented reality surgical guidance system for procedures involving a stereoscopic surgical microscope. Firstly, we show the sensitivity of a photogrammetric calibration method towards the detection of 2D markers in the projected calibration image and the markers' positional accuracy on the calibration object, both of them affecting the accuracy of the focal length estimate. Secondly, we propose a hybrid method using a photogrammetric method for pre-calibration and a number of self-calibration based methods to further optimise the focal length estimate. The results of the reported experiment, evaluating the proposed method, show a significant improvement in the calibration error of around 0.2 pixels as compared to calibrating each camera separately using a photogrammetric method only.

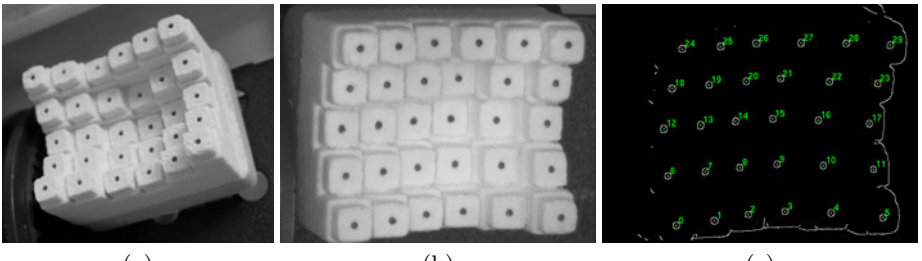
1 Introduction

Augmented reality (AR) based surgical guidance systems for minimally invasive surgery aim to combine pre- and intra-operatively acquired images. The intra-operatively acquired images ('real' imagery) are typically obtained from endoscopes or surgical microscopes, whereas the overlaying images ('virtual' imagery) are obtained from medical images (e.g. CT or MRI models). The first procedure in AR based surgical guidance is to determine the real camera system parameters as to align real and virtual images. This procedure is denominated calibration and involves estimating six extrinsic and five intrinsic camera parameters. The extrinsic parameters comprise six spatial degrees of freedom (DOF) in 3D, i.e. three for translation and three for rotation. The camera intrinsic parameters include the focal length, a scale factor, two coordinates of the principal point and a radial distortion parameter. A photogrammetric calibration method, e.g. Tsai's method [1], is typically used in conjunction with a planar or non-coplanar calibration object. This method determines the intrinsic and extrinsic parameters at separate stages during an iterative process, where the last stage determines the focal length. Tsai's method works well for estimating most of the calibration parameters. However, in specific circumstances, the focal length may present significant variations for small differences in input variables. The alternative is to use a self-calibrating method which does not strictly need a

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 551–560, 2010.

pre-defined calibration object in order to compute the camera parameters [2,3]. Instead, the procedure involves determining a variety of objects' features found in the real scene, which may include edges, corners, and regions of interest, among others.

Previous work in camera calibration for surgical microscopes has been presented by Edwards et al. [4] and Caversaccio et al. [5], among others. Their work is similar to ours regarding the estimation of camera parameters using photogrammetric methods, which requires an independent calibration for each of the cameras. Recent calibration methods aimed at optical devices with high distortion factors (i.e. endoscopes) have recently appeared in the literature [6]. Other approaches based on single cameras, like PTAM [7], estimate external camera parameters over time using scene features and can be used for medical or general-purpose AR based systems. This method requires initial camera parameters from a photogrammetric technique.



(a) (b) (c)

Fig. 1. (a) Manufactured 3D non-coplanar calibration object using the Z Corporation ZPrinter©450. (b) Top view of calibration object during calibration. (c) Detection of elliptical centres, numbered for easy identification.

In this paper, we present a hybrid calibration method to calibrate an AR surgical guidance system for head and neck surgery using a stereoscopic surgical microscope. The method uses a photogrammetric technique (Tsai) for pre-calibration and a self-calibrating stereoscopic technique to derive the fundamental matrix with the aim of optimising the calculated focal length. Self-calibration could be considered as less stable than photogrammetric methods due to the automatic selection of points in the scene, which usually generates a significant number of outliers. However, the use of a calibration object with known features provides higher point detection accuracy. In order to perform both pre-calibration (photogrammetric) and focal length optimisation (self-calibration), a planar calibration object is inadequate, hence we developed a 3D non-coplanar calibration object for this purpose - Figure 1(a)(b) - using a 3D printer (Z Corporation ZPrinter©450). The marker centres within the object are automatically detected [8] - Figure 1(c). In some instances projective distortions can affect the detection of circular markers if the disks are not parallel to the image, causing that the physical and projected circular/elliptical centres no longer coincide. Nevertheless, it has been proved that the coordinates of elliptical centres can be

corrected to correspond to disk circle centres even if the disks are not orthogonal to the image plane [9]. This correction can be done during the iterative calibration process. In the remainder of the paper, we first cover the photogrammetric method and illustrate its sensitivity to small deviations in marker centre detection and 3D marker position. Next, we introduce the hybrid method and show how it enables further optimisation of the calculated focal length.

2 Experiment I - Photogrammetric Calibration

The purpose of this section is to evaluate how the accuracy of 2D image marker detection and 3D marker position affects the camera calibration process using the manufactured calibration object in a magnified microscopic view. The experiments were carried out using a single camera in order to evaluate the stability of Tsai's photogrammetric algorithm [1], which produces the initial camera parameters to be optimised. These tests also serve to analyse the influence of the rotational position of the calibration object for the estimation of focal length, and consequently, the corresponding camera calibration errors. For this experiment a black and white camera was connected to one of the eyepieces of a stereoscopic surgical microscope, positioned perpendicularly to a flat bench. The non-coplanar calibration object (Figure 1) was placed on a rotating gauge that allows measuring the positional orientation at different angles with respect to the bench surface (which in turn is parallel to the camera image plane). This rotational instrument was attached to a height gauge to control the translational distance T_z between the calibration object and the microscope lens. A set of ten camera calibrations were performed for each slope angle, which varied from 0° to 25° at 5° steps. Inclination angles larger than 25° were excluded because at those orientations calibration markers went out of focus, affecting the localisation of circular shapes. These experiments extend our previous work [10] using endoscopic systems, where it was shown that calibration accuracy improves with increasing angle of inclination (also discussed by Tsai [1] and Zhang [11]). Table 1 shows the estimated focal length, f, translation in z, T_z, and the ratio of the former with the latter for each inclination angular position.

In addition, the remaining five DOFs were evaluated with the purpose of determining possible variations in the computed parameters at each inclination

Table 1. Mean and standard deviation (±SD) values for ten trials of focal length f, T_z distance and ratio between focal length and T_z at different inclination angles

Angle (degs)	Focal length f (mm)	T_z	f/T_z
0	202.2 (±4.2)	178.5 (±3.4)	1.133 (±0.003)
5	228.2 (±6.1)	199.1 (±5.1)	1.146 (±0.001)
10	265.9 (±10.7)	229.2 (±8.9)	1.160 (±0.001)
15	283.9 (±11.0)	242.6 (±9.3)	1.170 (±0.002)
20	303.6 (±6.4)	258.1 (±5.4)	1.176 (±0.000)
25	317.4 (±5.0)	268.9 (±4.2)	1.180 (±0.000)

angle. Table 2 demonstrates that the values obtained are consistent among all slope angles for parameters T_x, T_y, R_x and R_z. In the case of the rotational parameter R_y, the estimated values correspond to each inclination angle applied to the calibration object. A comparison between Table 1 and 2 shows that focal length is the most affected parameter during calibration.

Table 2. Mean and standard deviation (±SD) values for five different DOFs in a single camera calibration at different inclination angles over ten trials

Angle (degs)	$\boldsymbol{T_x}$	$\boldsymbol{T_y}$	$\boldsymbol{R_x}$	$\boldsymbol{R_y}$	$\boldsymbol{R_z}$
0	-10.25 (±0.02)	8.72 (±0.01)	173.05 (±0.10)	0.84 (±0.19)	-0.56 (±0.02)
5	-9.97 (±0.05)	8.69 (±0.02)	172.90 (±0.17)	5.65 (±0.50)	-0.70 (±0.04)
10	-10.26 (±0.00)	8.96 (±0.00)	172.60 (±0.03)	10.99 (±0.05)	-0.47 (±0.01)
15	-11.14 (±0.00)	9.23 (±0.00)	172.45 (±0.06)	16.31 (±0.06)	0.52 (±0.01)
20	-10.72 (±0.01)	9.30 (±0.01)	172.30 (±0.07)	21.13 (±0.19)	0.51 (±0.02)
25	-10.75 (±0.00)	8.95 (±0.00)	172.14 (±0.04)	24.68 (±0.04)	0.73 (±0.01)

Subsequently, we investigated the influence of 2D marker localisation inaccuracies on camera calibration errors. The set of projected marker points were subjected to Gaussian noise with 0 mean and three different standard deviation levels: 1.0, 0.5 and 0.3 pixels. For each noise level, ten independent calibrations were carried out at the same inclination angles as the above-mentioned tests. The results were averaged and compared to a ground truth obtained from an initial calibration at each orientation. The top row of Figure 2 shows absolute errors on the focal length for each noise level (left) and the ratio f/T_z (right). It is clear that the misidentification of marker centres to the level of just half a pixel in a microscopically-magnified view has a considerable effect on the accuracy of the focal length, and even when the former is compensated by T_z through the f/T_z ratio. Heikkilä and Silvén [9] also pointed out the sensitivity of camera parameters to 2D marker localisation.

Finally, we investigated the effect of positional accuracy of the 3D marker points - largely dependent on the accuracy by which the calibration object is manufactured. Gaussian noise with 0 mean and standard deviations of 0.3, 0.1 and 0.05mm were added to the nominal coordinates of the constructed model. The results are shown in the bottom row of Figure 2. It is clear that the effect on the focal length and f/T_z is even larger than for 2D marker localisation. Considering that the 3D printer's accuracy is of the level of 0.05-0.1mm, either higher accuracies are needed to manufacture the object or an alternative optimisation could be employed, which we discuss in the following section.

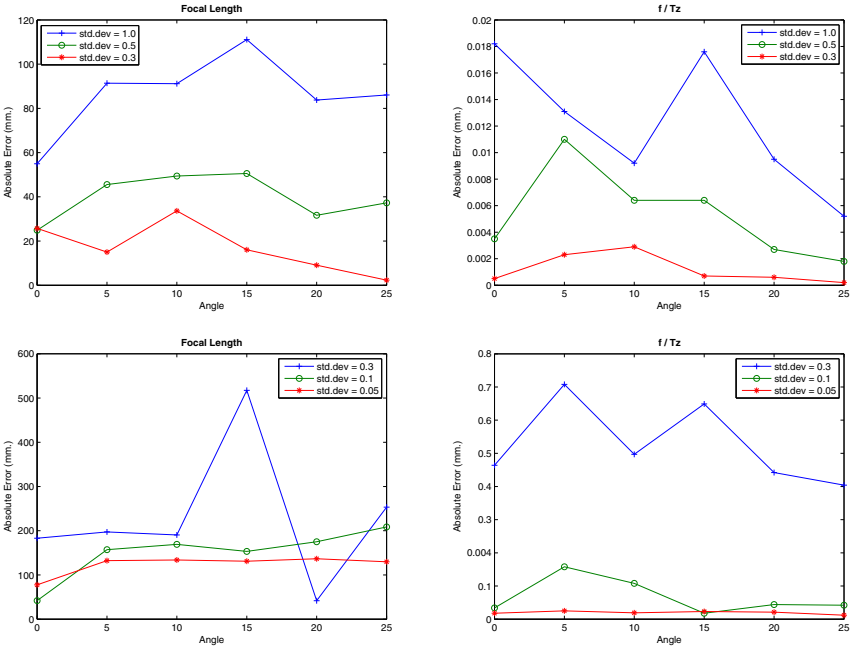


Fig. 2. Top row: effect of marker's 2D localisation inaccuracies - Left: Focal length absolute error; Right: f/T_z absolute error. Bottom row: effect of 3D marker positional accuracy - Left: Focal length absolute error; Right: f/T_z absolute error. The abscissa shows the inclination angle of the calibration object in degrees.

3 Experiment II: Hybrid Calibration

When using a photogrammetric method only, each camera of the stereo pair is calibrated separately. Here, we will exploit the stereopsis property by calculating the fundamental matrix [12] and subsequent focal length optimisation for our real-time stereo camera assembly mounted on the surgical microscope. Initial focal length estimations for each of the left and right images were obtained by using the earlier discussed photogrammetric method (Tsai). The setup involved a stereoscopic surgical microscope placed on a flat work bench and two cameras connected to the microscope eyepieces, which are initially placed in a parallel position. It is worth mentioning that this setup would make the self-calibration method fail because the relative alignment between cameras belongs to a degenerate configuration [13]. In order to avoid this critical configuration, it was decided to acquire the left camera image at a still pose while changing both convergence (angle inclination) and elevation on the right view. The approach is analogous to other photogrammetric calibration methods that require changing the orientation of the camera or calibration object between at least two different views [11]. Ten independent focal length estimations were executed for each

convergence and elevation positions, within a range from 0° to 30° in the case of convergence and 0° to 15° for elevation, both at 5° steps. Higher inclination levels affected the detection of elliptical centres and were excluded from the evaluation. The overall optimisation procedure involves three steps, each containing three optional algorithms:

1. The first step is the calculation of the fundamental matrix, F, from at least seven corresponding points on the calibration object, in the left and right images. Three different algorithms are considered: linear [14], gradient-based and the M-estimators [15,16].
2. Next, the focal length, f, is calculated from the previous estimated fundamental matrix. Three self-calibration based methods, i.e. Bougnoux [17], Sturm [13,18] and Newsam [19] are considered. The calculation of f from Tsai's photogrammetric method was used as the ground truth.
3. One of the limitations of estimating the focal length through self-calibration is that, in order to recover a reliable solution, the effects of radial distortion for each of the cameras must be initially corrected [20]. Subsequently, the pair of focal lengths computed by any of the methods described in Step 2 can be refined through an optimisation algorithm. As the initial extrinsic and intrinsic camera parameters have already been estimated at the pre-calibrated stage, such knowledge can be included in a cost function in order to improve the solution. The cost function that has been selected relies on a metric known as the Sampson distance [21], which is a first-order approximation to a geometric, or reprojection, error measured in left and right images. The cost function is defined as:

$$\sum_i \frac{(\hat{p}_i^T E p_i)^2}{(Ep_i)_1^2 + (Ep_i)_2^2 + (E^T\hat{p}_i)_1^2 + (E^T\hat{p}_i)_2^2}, \tag{1}$$

   where E is the essential matrix obtained from Equation 2:

$$E = \hat{K}^T F K, \tag{2}$$

   with fundamental matrix, F, and the pair of pre-calibrated intrinsic camera matrices K and $\hat{K}$ for left and right images, respectively. p and $\hat{p}$ in Equation 1 represent corresponding image points m and $\hat{m}$ in normalised coordinates [21], which are obtained from $p = K^{-1}m$ and $\hat{p} = \hat{K}^{-1}\hat{m}$, respectively. Therefore, the parameters to optimise involve the intrinsic camera matrices with respect to the pair of focal length values.
   The three different optimisation methods include two evolutionary algorithms, Self-adaptive Differential Evolution (SDE) [22] and CODEQ [23], and the well-established Levenberg-Marquardt algorithm.

Thus, a complete evaluation comprised 27 combinations among the fundamental matrix and focal length techniques for each optimisation algorithm. Because self-calibration methods do not provide information about accuracy errors by themselves, the refined focal lengths were fed back into the original Tsai's method and

a recomputation of camera calibration on both cameras was carried out (maintaining the other extrinsic and intrinsic pre-calibrated parameters). In other words, Tsai's method was performed an additional time in order to calculate the calibration error given the refined focal lengths. This provided a means to compare the calibration accuracy with respect to the initial (ground truth) calibration errors. The parameters used for the optimisation involved a maximum number of 400 iterations for the Levenberg-Marquardt algorithm, whereas a maximum of 20 generations with a population size of 20 individuals was selected for both evolutionary algorithms (SDE, CODEQ). The tolerance threshold in the cost function (Equation 1) comprised a value of 10^{-16} for all optimisation methods based on an initial function value of 1.8^{-7}. In general, the time taken for the three optimisation algorithms to refine the focal length parameter was a couple of seconds on a modern PC with dual core processor. The Levenberg-Marquardt optimisation method produced unstable results as compared to the genetic algorithm optimisation methods and is therefore not shown in Figure 3, which shows plots of the calibration error in pixels for the remaining 18 combinations at an elevation angle of 15°. Among the different elevations tested, an angle of 15° was the most favourable case, where the absolute difference between Tsai's ground truth and mean image calibration errors is more noticeable. Therefore, the reported analysis of these experiments is focused at this elevation for the diverse fundamental matrix, stereo focal length and optimisation methods.

4 Discussion

The first important observation from Figure 3 is that the calibration error after applying the hybrid method described in section 3 for all combinations shows significant improvement as compared to the initial calibration using Tsai's method only (ground truth). It can also be seen that there is little difference between the SDE and CODEQ methods. Among the three algorithms used to compute the fundamental matrix, the gradient-based method provided slightly better results than the linear counterpart and proved to be more stable than the M-estimators method. The results obtained by the different methods of Bougnoux, Sturm and Newsam show similar performance when using the gradient method. However, all of these are statistically insignificant differences as compared to the substantial improvement of around 0.2 pixels, after using the hybrid method (for more information, the reader is referred to [8]). Note that there is a break in the curves representing the methods of Sturm and Newsam at 25° of inclination (convergence) angle, which means that both techniques failed to estimate a focal length. Although this is not caused by a critical configuration (e.g. parallel camera setup), it is assumed that this is produced by certain instabilities within these algorithms.

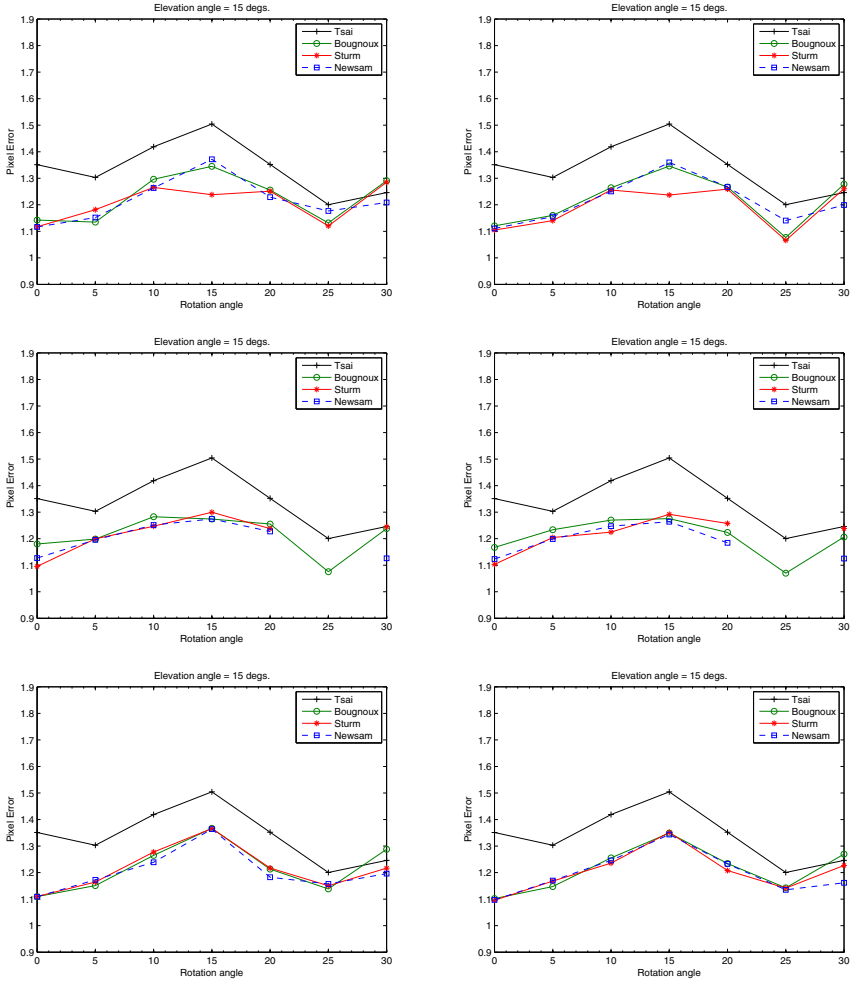


Fig. 3. Left column: Self-adaptive Differential Evolution optimisation; Right Column: CODEQ optimisation. Top row: Linear; Middle row: M-estimator; Bottom row: Gradient based - fundamental matrix estimation. Each of the three algorithms to calculate the focal length from the fundamental matrix are shown on each plot with Tsai's algorithm as the ground truth.

5 Conclusion

This paper focused on the study of camera calibration for the calculation of intrinsic and extrinsic parameters, which is required as an initial stage to superimpose virtual imagery in an AR based surgical guidance scene. Firstly, an analysis of Tsai's photogrammetric calibration in a single camera was performed using a non-coplanar calibration object. It was demonstrated that the stabil-

ity of the algorithm is significantly influenced by the precision of the object's physical construction and the detection of image markers, especially for magnified imagery acquired by a surgical microscope. Most importantly, it was shown that focal length is the most affected parameter with respect to external conditions, the orientation of the calibration device being one of them. Secondly, an evaluation of different methods for the estimation of focal length was carried out. The purpose was to optimise the focal length and consequently improve the accuracy of the final calibration error. The results indicate that the use of an evolutionary algorithm can decrease the original calibration errors obtained by the photogrammetric method when one of the cameras is rotated about 15° around the elevation axis. Specifically, the best combination comprised the use of the gradient-based method for the fundamental matrix and CODEQ optimisation, where the three techniques for the computation of focal length (i.e. Bougnoux, Newsam and Sturm) produced similar results. Although in previous work [10] we have obtained sub-pixel accuracy using a planar calibration object through Tsai's method (0.3 - 0.4 pixels), the presented non-coplanar calibration object was less accurately manufactured due to limitations of the 3D printer employed. We expect that by improving the precision of the non-coplanar calibration object the same level of initial calibration accuracy will be obtained which can be further minimised by applying the proposed hybrid method. It should be noted that the presented technique can be extended to non-stereoscopic devices such as endoscopes where the endoscope is placed at different positions to obtain multiple image acquisitions. Of course, this would require the additional use of a tracking device which, depending on its accuracy, may affect the final optimisation.

References

1. Tsai, R.: A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. IEEE Journal of Robotics and Automation 3(4), 323–344 (1987)
2. Faugeras, O., Luong, Q.T., Maybank, S.: Camera self-calibration: Theory and experiments. In: Sandini, G. (ed.) ECCV 1992. LNCS, vol. 588, pp. 321–334. Springer, Heidelberg (1992)
3. Lourakis, M., Deriche, R.: Camera self-calibration using the singular value decomposition of the fundamental matrix. In: 4th Asian Conference on Computer Vision, vol. 1, pp. 403–408 (2000)
4. Edwards, P., King, A., Maurer, C., de Cunha, D., Hawkes, D., Hill, D., Gaston, R., Fenlon, M., Chandra, S., Strong, A., Chandler, C., Richards, A., Gleeson, M.: Design and evaluation of a system for microscope-assisted guided interventions (MAGI). IEEE Transactions on Medical Imaging 19, 1082–1093 (2000)
5. Caversaccio, M., Garcia-Giraldez, J., Gonzalez-Ballester, M., Marti, G.: Image-guided surgical microscope with mounted minitracker. The Journal of Laryngology & Otology 121, 160–162 (2007)
6. Barreto, J., Roquette, J., Sturm, P., Fonseca, F.: Automatic camera calibration applied to medical endoscopy. In: British Machine Vision Conference, BMVC 2009 (2009)

7. Klein, G., Murray, D.: Parallel tracking and mapping for small AR workspaces. In: International Symposium on Mixed and Augmented Reality, ISMAR 2007 (2007)
8. González-García, G.: Optimised Calibration, Registration and Tracking for Image Enhanced Surgical Navigation. PhD thesis, University of East Anglia, United Kingdom (March 2010)
9. Heikkilä, J., Silvén, O.: A four-step camera calibration procedure with implicit image correction. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 1997), pp. 1106–1112 (1997)
10. Lapeer, R., Chen, M., Gonzalez, G., Linney, A., Alusi, G.: Image-enhanced surgical navigation for endoscopic sinus surgery: Evaluating calibration, registration and tracking. The International Journal of Medical Robotics and Computer Assisted Surgery (IJMRCAS) 4(1), 32–45 (2008)
11. Zhang, Z.: Flexible camera calibration by viewing a plane from unknown orientations. In: International Conference on Computer Vision (ICCV 1999), vol. 1, pp. 666–673 (September 1999)
12. Luong, Q.: Fundamental Matrix and Self-calibration. PhD thesis, University of Paris, Orsay (December 1992)
13. Sturm, P., Cheng, Z., Chen, P., Poo, A.: Focal length calibration from two views: Method and analysis of singular cases. Computer Vision and Image Understanding 99(1), 58–95 (2005)
14. Longuet-Higgins, H.: A computer algorithm for reconstructing a scene from two projections. Nature 293, 133–135 (1981)
15. Armangué, X., Salvi, J.: Overall view regarding fundamental matrix estimation. In: Image and Vision Computing, pp. 205–220 (2003)
16. Zhang, Z.: Determining the epipolar geometry and its uncertainty: A review. International Journal of Computer Vision 27(2), 161–198 (1998)
17. Bougnoux, S.: From projective to euclidean space under any practical situation, a criticism of self-calibration. In: International Conference on Computer Vision (ICCV 1998), pp. 790–796 (1998)
18. Sturm, P.: On focal length calibration from two views. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001) 2, 145–150 (2001)
19. Newsam, G., Huynh, D., Brooks, M., Pan, H.P.: Recovering unknown focal lengths in self-calibration: An essentially linear algorithm and degenerate configurations. Int. Arch. Photogrammetry Remote Sensing 31(B3), 575–580 (1996)
20. Tordoff, B., Murray, D.: Violating rotating camera geometry: The effect of radial distortion on self-calibration. In: 15th International Conference on Pattern Recognition (ICPR 2000), vol. 1, pp. 423–427 (2000)
21. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge (2003)
22. Salman, A., Engelbrecht, A., Omran, M.: Empirical analysis of self-adaptive differential evolution. European Journal of Operational Research 183(2), 785–804 (2007)
23. Omran, M., Salman, A.: Constrained optimization using CODEQ. Chaos, Solitons and Fractals 42(2), 662–668 (2009)

An Efficient Graph-Based Deformable 2D/3D Registration Algorithm with Applications for Abdominal Aortic Aneurysm Interventions

Rui Liao[1], Yunhao Tan[2], Hari Sundar[1], Marcus Pfister[3], and Ali Kamen[1]

[1] Imaging and Visualization Department,
Siemens Corporate Research, USA
{rui.liao,hari.sundar,ali.kamen}@siemens.com
[2] Department of Computer Science,
Wayne State University, USA
tanyunhao@wayne.edu
[3] Healthcare Sector, Siemens AG, Forchheim, Germany
marcus.pfister@siemens.com

Abstract. 2D/3D registration is in general a challenging task due to its ill-posed nature. It becomes even more difficult when deformation between the 3D volume and 2D images needs to be recovered. This paper presents an automatic, accurate and efficient deformable 2D/3D registration method that is formulated on a 3D graph and applied for abdominal aortic aneurysm (AAA) interventions. The proposed method takes the 3D graph generated from a segmentation of the CT volume and the 2D distance map calculated from the 2D X-ray image as the input. The similarity measure consists of a difference measure, a length preservation term and a smoothness regularization term, all of which are defined and efficiently calculated on the graph. A hierarchical registration scheme is further designed specific to the anatomy of abdominal aorta and typical deformations observed during AAA cases. The method was validated using both phantom and clinical datasets, and achieved an average error of $< 1mm$ within $0.1s$. The proposed method is of general form and has the potential to be applied for a wide range of applications requiring efficient 2D/3D registration of vascular structures.

1 Introduction and Background

Abdominal aortic aneurysm (AAA) is the local expansion of the abdominal aorta. There is a risk of rupture of the aneurysm if the expansion becomes large enough, and the chances of post-rupture survival for the patients are low. AAA is currently ranked as the 13^{th} leading cause of death in the U.S. [1]. In recent years, minimal-invasive interventional repairs are rapidly emerging as an alternative to open surgeries for the treatment of AAA, especially for patients at an increased surgical risk due to age or other medical conditions. During the interventional procedure, X-ray imaging is routinely used for the guidance and navigation of the catheter and graft within the aorta. Overlay of the pre-operative 3D volumetric data onto the intra-operative 2D X-ray images can provide realistic artery

H. Liao et al. (Eds.): MIAR 2010, LNCS 6326, pp. 561–570, 2010.

anatomy and introduce useful information to the physicians for finding the best path and target position. Due to the deformable nature of the abdominal organs, there typically exists elastic deformations between pre-operative volumes and intra-operative X-ray images. The insertion of the medical devices into the artery during AAA procedures can further introduce significant deformations to the target vessel that needs to be aligned. Although several methods have been proposed for deformable 2D/3D registration [2,3,4], they are not tailored for vascular structures. In [5], a deformable method is proposed for registering 3D vessel structures to a single projection image. However, the method in [5] formulated the registration problem inadequately and ended up with a very inefficient solution. In particular, 2D and 3D centerlines are generated from the corresponding vessel segmentations, and correspondences are estimated together with the deformation field via iterative optimization. For Ni points for a given iD centerline ($i = 2, 3$), we need to estimate $N2 \times N3$ parameters for the correspondence, and $N3 \times 3$ parameters for the deformation field. This is challenging both in terms of computational complexity and numerical stability. Thin-plate-splines (TPS) are used to enforce smoothness, which further increases the computational complexity. A 5-min run-time is reported in [5] for their registration step, making the algorithm impractical to use during interventional procedures such as AAA.

In this paper, we propose to reformulate the registration problem on a 3D graph, with improvements on all three terms in the similarity measure used in [5]. We further provide an efficient numerical solution to the graph-based formulation. In particular, a 3D graph is generated from the abdominal aorta segmented from the pre-operative CT data, and a 2D distance map is generated from each of the 2D X-ray images used for registration. A distance map is a smooth shape encoding of the underlying structures, and has been applied to various registration problems [6,7]. By utilizing a distance map, explicit establishment of point correspondences between 2D and 3D graphs can be avoided during the optimization. This reduces the optimization space to a much lower dimension of $N3 \times 3$. In addition, smoothness calculation is defined on the 3D graph, the derivative of which can be calculated efficiently using the well-known Laplacian matrix of a graph. Specific to the anatomy of abdominal aorta, a hierarchical registration scheme is further deployed. In particular, the 3D graph is divided into three segments, renal arteries, iliac arteries, and abdominal aorta. A piecewise rigid-body transformation is first applied individually to the three segments while their connectivity is maintained. Local deformation is then estimated for the complete graph comprising all the three segments. In contrast to the method in [5], which uses a single projection image for registration, we further compare the registration accuracy achieved by using one and two views. We show that even with the incorporation of length-preserving term, a single projection image alone can only produce accurate registration in the imaging plane, while two views are required in order to achieve accurate registration in 3D.

2 Method

The workflow of our proposed method is depicted in Fig. 1. First, the abdominal aorta is segmented from CT volumes using a graph-cut based segmentation method [8]. The centerlines are generated from the segmentation masks by a sequential topological thinning processes [9], and 3D graphs are subsequently generated from the extracted centerlines. Second, vessel system in 2D X-ray images is segmented and a 2D distance map is generated [10]. A distance map provides an efficient way of computing the distance between two centerlines, and needs to be computed only once before the iterative optimization starts. In comparison, the method in [5] finds the closest point pair iteratively by estimating the correspondences explicitly and that significantly increases the number of parameters to be optimized. The above steps for preprocessing on both 2D and 3D data have been previously presented in the literatures and are not within the scope of the paper. In addition, segmentation in AAA cases is relatively straightforward for a contrasted 3D volume and 2D DSA images with negligible motion compared to cardiac applications. Using these given inputs, we define our energy functional. Finally, the 3D graph is registered to the 2D distance map by the following variational approach, in which all three terms are formulated on a 3D graph.

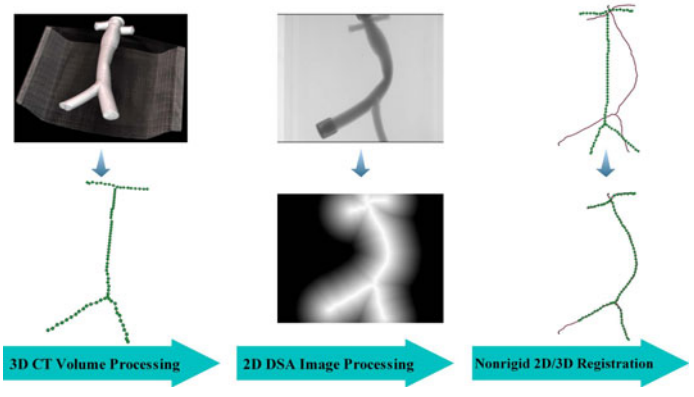


Fig. 1. The pipeline of deformable 2D/3D registration for AAA interventions. Note that, in all the illustrations through this paper, 3D graph features green nodes while ground truth features red edges.

Our 3D graph G is an undirected, acyclic, and single-component graph, with n nodes $\mathbf{x_{set}} = \{\mathbf{x_i}|\ i = 1...n\}$, and m edges of length $\mathbf{l_{set}} = \{l_i|\ i = 1...n\}$. For each node $\mathbf{x}_i \in \mathbf{x}$, its first ring neighbor edge is denoted by e_i, where e_i contains t_i number of edges. The energy function to be minimized is:

$$\mathcal{E}(\mathbf{u}) = \mathcal{D}(\mathbf{u}) + \alpha\mathcal{S}_{\mathcal{L}}(\mathbf{u}) + \beta\mathcal{S}_{\mathcal{D}}(\mathbf{u}), \tag{1}$$

where $\mathbf{u_{set}} = \{\mathbf{u_i}|\ i = 1...n\}$ represents the node displacement to be estimated by minimizing the above energy function. The final position of the node $\mathbf{x}_i$ therefore is $\mathbf{y}_i = \mathbf{x}_i + \mathbf{u}_i$. In the energy functional, $\mathcal{D}$ is the image-based difference

measure, $\mathcal{S}_\mathcal{L}$ is the length preserving term and $\mathcal{S}_\mathcal{D}$ is the smoothness regularization term. α and β are the coefficients balancing the corresponding terms in the energy functional. We adopt a gradient-based optimizer to gain computational efficiency. For this reason we need to compute the derivative of each term in Eq. 1 with respect to node displacement $\mathbf{u}_i$.

Difference Measure. The difference measure $\mathcal{D}$ is computed as follows:

$$\mathcal{D} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{M}^2(\mathbf{d}(\mathbf{y}_i)), \tag{2}$$

where $\mathbf{d}(\mathbf{y}_i)$ is the 2D projection of i^{th} displaced node $\mathbf{y}_i$ and $\mathbf{M}$ is the 2D distance map. Let $\mathbf{y}_i^h$ denote the homogeneous coordinate of the i^{th} node, i.e., $\mathbf{y}_i^h = \begin{bmatrix}\mathbf{y}_i^\top & 1\end{bmatrix}^\top$, therefore, $\mathbf{d}(\mathbf{y}_i) = \begin{bmatrix}\frac{\mathbf{p}_1^\top \mathbf{y}_i^h}{\mathbf{p}_3^\top \mathbf{y}_i^h} & \frac{\mathbf{p}_2^\top \mathbf{y}_i^h}{\mathbf{p}_3^\top \mathbf{y}_i^h}\end{bmatrix}^\top$, where the 3×4 projection matrix $\mathbf{P} = \begin{bmatrix}\mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3\end{bmatrix}^\top$. Derivative of $\mathcal{D}$ with respect to $\mathbf{u}_i$ is

$$\frac{\partial \mathcal{D}}{\partial \mathbf{u}_i} = \frac{1}{(\mathbf{p}_3 \mathbf{y}_i^h)^2} \begin{bmatrix} \mathbf{p}_1^{*\top}(\mathbf{p}_3^\top \mathbf{y}_i^h) - \mathbf{p}_3^{*\top}(\mathbf{p}_1^\top \mathbf{y}_i^h) \\ \mathbf{p}_2^{*\top}(\mathbf{p}_3^\top \mathbf{y}_i^h) - \mathbf{p}_1^{*\top}(\mathbf{p}_2^\top \mathbf{y}_i^h) \end{bmatrix}^\top \nabla \mathbf{M}, \tag{3}$$

where $*$ denotes that only the first three elements of the vector is considered. It can be seen that the gradient of the distance map ($\nabla \mathbf{M}$) needs to be calculated only once and is used for efficient gradient calculation in Eq. 3 by proper parametrization of the gradient. The distance measure serves as a three dimensional force that moves each node toward the direction, where its projection onto the distance map image is minimized.

Length Preserving Term. We assume that movement of the nodes are constrained in the way the total length of the vessel is preserved. A stronger constraint is that length of each edge in the graph is preserved. We formulate the edge length preserving term $\mathcal{S}_\mathcal{L}$ on the 3D graph as follows:

$$\mathcal{S}_\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{t_i} \left(\frac{l_j^p - l_j}{l_j}\right)^2, \tag{4}$$

where l_j is the original length of an edge, and l_j^p is the new length of the same edge after the deformation, i.e., $l_j^p = \|(\mathbf{x}_j + \mathbf{u}_j) - (\mathbf{x}_i + \mathbf{u}_i)\|$. The derivative of $\mathcal{S}_\mathcal{L}$ with respect to $\mathbf{u}_i$ is computed as follows:

$$\frac{\partial \mathcal{S}_\mathcal{L}}{\partial \mathbf{u}_i} = \frac{1}{n}\sum_{j=1}^{t_i} \frac{\partial (\frac{l_j^p - l_j}{l_j})^2}{\partial \mathbf{u}_i} = \frac{1}{n}\sum_{j=1}^{t_i} \frac{2(l_j^p - l_j)}{(l_j)^2 l_j^p} \left((\mathbf{x}_i + \mathbf{u}_i) - (\mathbf{x}_j + \mathbf{u}_j)\right). \tag{5}$$

When $l_j^p < l_j$, this term forces the node to move away from its j^{th} first-ring neighbor, and vice versa.

Smoothness Term. We also assume that neighboring nodes on the graph move coherently. The smoothness term, $\mathcal{S}_{\mathcal{D}}$, is computed as follows:

$$\mathcal{S}_{\mathcal{D}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{t_i}\|\mathbf{u}_i - \mathbf{u}_j\|^2. \tag{6}$$

The derivative of $\mathcal{S}_{\mathcal{D}}$ with respect to $\mathbf{u}_i$ is

$$\frac{\partial \mathcal{S}_{\mathcal{D}}}{\partial \mathbf{u}_i} = \frac{1}{n}\sum_{j=1}^{t_i}\frac{\partial\|\mathbf{u}_i - \mathbf{u}_j\|^2}{\partial \mathbf{u}_i} = \frac{1}{n}\sum_{j=1}^{t_i} 2(\mathbf{u}_i - \mathbf{u}_j). \tag{7}$$

This term forces the node to move toward the barycenter of its first ring neighbors, and can be calculated efficiently using the Laplacian matrix of a graph.

Optimization Scheme. Since this is an unconstrained nonlinear optimization problem, we considered using BFGS method [11] to minimize the energy functional of Eq. 1. Table 1 outlines the steps of the method, where parenthesized superscript denotes intermediate results for the k^{th} iteration.

Table 1. Deformable 2D/3D registration algorithm flowchart

Algorithm Flowchart:

repeat

1. Obtain a direction $\mathbf{v}^{(k)}$ by solving: $\mathbf{v}^{(k)} = -\mathbf{B}^{(k)^{-1}}\nabla\mathcal{E}(\mathbf{u}^{(k)})$;
2. Perform a line search to find the best step size, $\lambda^{(k)}$; then update $\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \lambda^{(k)}\mathbf{v}^{(k)}$;
3. Let $\mathbf{u}^{(k)} = \lambda^{(k)}\mathbf{v}^{(k)}$;
4. Calculate the gradient difference, $\mathbf{e}^{(k)} = \nabla\mathcal{E}(\mathbf{u}^{(k+1)}) - \nabla\mathcal{E}(\mathbf{u}^{(k)})$;
5. Calculate the Hessian matrix, $\mathbf{B}^{(k+1)} = \mathbf{B}^{(k)} + \frac{\mathbf{e}^{(k)}\mathbf{e}^{(k)\top}}{\mathbf{e}^{(k)\top}\mathbf{u}^{(k)}} - \frac{\mathbf{B}^{(k)}\mathbf{u}^{(k)}(\mathbf{B}^{(k)}\mathbf{u}^{(k)})^{\top}}{\mathbf{u}^{(k)\top}\mathbf{B}^{(k)}\mathbf{u}^{(k)}}$;
6. Compute $\mathbf{B}^{(k+1)^{-1}}$ from 5 or use Sherman-Morrison formula [12]:

$$\mathbf{B}^{(k+1)^{-1}} = \mathbf{B}^{(k)^{-1}} + \frac{(\mathbf{u}^{(k)\top}\mathbf{e}^{(k)} + \mathbf{e}^{(k)\top}\mathbf{B}^{(k)^{-1}}\mathbf{e}^{(k)})(\mathbf{u}^{(k)}\mathbf{u}^{(k)\top})}{(\mathbf{u}^{(k)\top}\mathbf{e}^{(k)})^2} - \frac{\mathbf{B}^{(k)^{-1}}\mathbf{e}^{(k)}\mathbf{u}^{(k)\top} + \mathbf{u}^{(k)}\mathbf{e}^{(k)\top}\mathbf{B}^{(k)^{-1}}}{\mathbf{u}^{(k)\top}\mathbf{e}^{(k)}};$$

until $\|\mathbf{e}^{(k)}\| < \varepsilon$;

The registration is performed in a hierarchical manner. We first estimate the global rigid alignment[1] between the two aorta by aligning the two bifurcation points (the illiac and renal bifurcations). The translation at the two bifurcation points is applied to the respective branches. This step yields us a good guess of $\mathbf{u}^{(0)}$. Following this, we optimize for the displacement at all nodes of the 3D graph. $\mathbf{B}^{(0)}$ is initialized with identity. The first iteration is then equivalent to a gradient descent, but further iterations will be accelerated by $\mathbf{B}^{(k)}$, which is the approximated Hessian matrix. For more details, readers are referred to [11].

[1] This step efficiently excludes the risk of BFGS falling into a local minimum.

3 Experiments and Results

In order to validate our proposed method, we conducted a series of experiments, which included the 3D volumes before and after deformation to establish ground truth. Using this data, we quantitatively measured the 3D target registration error (TRE) based on the 3D Euclidean distance between the transformed nodes of the graph and the corresponding ground truth, defined as $TRE = \frac{1}{n}\sum_{i=1}^{n} \|(\mathbf{x}_i+\mathbf{u}_i)-\mathbf{g}_i\|$, where $\mathbf{g}_i$ is the ground truth location of the node corresponding to $\mathbf{x}_i$. The mean value and standard deviation of the registration error are presented as μ and σ, respectively. In addition, re-projection error is calculated by the average of distance map lookups based on the projected 3D points as mentioned in Eq. 2. The calibration for the central projection matrix are done off-line and once at the beginning of the procedure for a given machine.

3.1 Phantom Data

A AAA phantom (top left of Fig. 1) comprising the simulated abdominal aorta, renal arteries, and iliac arteries was used for our phantom study. A set of DynaCT rotational run was acquired on a biplane C-arm system (AXIOM Artis, Siemens Medical Solutions, Erlangen, Germany) about the phantom that is filled with radio-opaque fluid. Artificial deformation was introduced to the phantom to simulate the deformation caused by device insertion and/or organ movement. A total of 9 datasets were acquired, resulting in a total number of $9 \times 8 = 72$ pairs for phantom validation. For a given pair of datasets, the 3D graph from the first dataset and the 2D distance map from the second dataset were used as the input for our registration algorithm, and the 3D centerline from second dataset was used as the ground truth for TRE calculation. One or two 2D X-ray images ($\sim$90 degrees apart) were used for each pair. Some examples are shown in Fig. 2 and the quantitative results comparing the offset before and after registration are summarized in Fig. 3 and Table 2. It can be seen that when two 2D X-ray images were used, the registration result was highly accurate in 3D physical space (Fig. 2-(d-e)). In comparison, a single projection image alone can produce accurate registration in the imaging plane, while the estimation in the depth direction

Table 2. Experimental results with two-view setting. Note that, to obtain the ground truth for the real patient data, synthesized and natural deformation are applied to the 3D graph. The results are presented in "Clinical Data with Synthetic Deformation" and "Clinical Data with Natural Deformation" respectively. α and β are 200 in these experiments.

Test Type	μ $Initial/Registered(mm)$	σ $Initial/Registered(mm)$	Runtime (s)
Phantom Data, All 72 Pairs	8.55/0.76	3.91/0.33	0.075
Clinical Data, with Synthetic Deformation	15.19/0.95	9.27/0.87	0.092
Clinical Data, with Natural Deformation	9.72/0.82	5.50/0.73	0.088

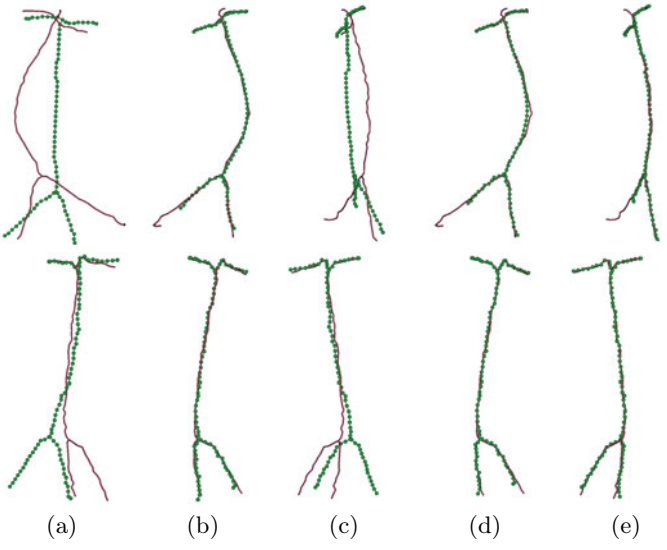


Fig. 2. Two examples of phantom experiments, one row for each case. (a) The 3D graph and the ground truth are overlaid to show their initial offset; (b-c) Registration results using one X-ray image. The overlay is accurate from the imaging plane (b), but not in the physical space as shown from a secondary angle (c); (d-e) Registration results using two X-ray images. The overlay is shown from the same angles as those for (b-c).

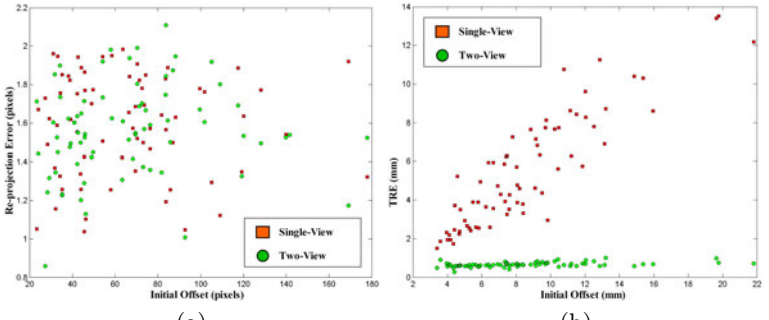


Fig. 3. Scatter plot of registration results for 72 pairs of phantom test. The registration TRE is plotted versus the initial offset. (a) Re-projection error in the 2D imaging plane, which is small across a wide range of initial offsets when either one- or two- views are used; (b) TRE in 3D, which is consistently small across a wide range of initial offsets when two X-ray images are used, demonstrating a relatively large capture range of the proposed hierarchical registration scheme.

can be largely off (Fig. 2-(b-c)). Compared to the method reported in [5] where registration took ∼5 mins, our registration algorithm took less than 0.1s with comparable accuracy. Both methods were implemented in C++ and tested on

Fig. 4. (a) An example of patient CT data, with AAA segmented and shown in mesh; (b) The 3D graph and the ground truth with synthesized deformation for patient one; (c) Registration results of (b); (d) The 3D graph and the ground truth with natural deformation for patient two; (e) Registration result of (d).

similar machine configuration. The significantly improved efficiency and satisfactory accuracy makes the propose algorithm highly suitable for interventional applications that require efficient update of registration during the procedures. In addition, by integrating the hierarchical registration scheme, our registration algorithm performed robustly across a wide range of initial offsets (see Fig. 3).

3.2 Clinical Data

Our proposed method was also applied on clinical data from two patients suffering from AAA. The ground truth for real patient data is difficult to obtain. In our experiments, the 3D graph was generated from a CT volume of the patient, and was deformed by two types of known deformations, one synthesized and one natural. The synthesized deformation field was generated using a length preserving deformation function (Fig. 4-(b)). The natural deformation field represented real clinical scenario and was obtained by registering the CT volume to the DynaCT volume acquired during the intervention for the respective patient (Fig. 4-(d)). The introduced offset was estimated to average at 15.19 mm and 9.75 mm for the synthesized and natural deformations respectively (Table 2). The deformed 3D graph using either synthesized or natural deformation fields was then used as the ground truth for quantifying the registration performance. A projection matrix computed from a rigid CTA-to-DSA registration of the respective patient is used to create the input 2D distance map. The proposed registration was then applied using the original 3D graph and the created 2D distance map as the input, and the registration result was compared to the ground truth. The advantage of using the simulated instead of real 2D projections is that we have the ground truth and hence can validate our registration accuracy in 3D, which is essential. Some experimental results are shown in Fig. 4 and quantitative results are summarized in Table 2.

4 Discussion and Conclusion

In this paper, we presented an accurate and efficient graph-based deformable 2D/3D registration method and applied it on AAA data. A 3D graph is generated to represent the vascular structure in 3D, and a 2D distance map is computed to smoothly encode the centerline of the vessel shape. This enables us to formulate the problem without a need for establishing explicit correspondence on a 3D graph. Two views were used for deriving an accurate registration in 3D, and a hierarchical registration scheme increased the capture range significantly. We applied BFGS optimizer by proper implementation of the gradient of the different terms of the energy functional explicitly on the 3D graph. We achieved sub-millimeter accuracy on both phantom and clinical data, which is comparable to that reported in [5]. The speed-up is over 1000 times. Efficiency is gold in interventional applications. Even though contrast is only administered from time to time, the physicians can not wait a long time for the registration result before resuming interventions. From our clinical partners, an acceptable registration time is $<$30s for initial registration and $<$5s for updates. The substantially improved efficiency further opens the door to potential applications. For example, the weight for each term in the energy functional is data-dependent. Our run-time of 0.1s makes it feasible to get the optimal combination of weights interactively for a given data, even during interventions. The method is not application specific and has the potential to be applied on a wider set of problems including cardiology and neuro interventional procedures. This is our future work.

References

1. Belkin, M., Donaldson, M.C., Whittemore, A.D.: Abdominal aortic aneurysms. Current Opinion in Cardiology 9(5), 581–590 (1994)
2. Bodensteiner, C., Darolti, C., Schweikard, A.: A fast intensity based non-rigid 2d-3d registration using statistical shape models with application in radiotherapy. In: Astola, J.T., Egiazarian, K.O., Nasrabadi, N.M., Rizvi, S.A. (eds.) Proceedings of the SPIE, vol. 7245, 72450G (2009)
3. Qi, W., Gu, L., Xu, J.: Non-rigid 2d-3d registration based on support vector regression estimated similarity metric. In: Dohi, T., Sakuma, I., Liao, H. (eds.) MIAR 2008. LNCS, vol. 5128, pp. 339–348. Springer, Heidelberg (2008)
4. Prűmmer, M., Hornegger, J., Pfister, M., Dőrfler, A.: Multimodal 2D-3D non-rigid registration. In: Reinhardt, J.M., Pluim, J.P.W. (eds.) Proceedings of the SPIE, vol. 6144, pp. 297–308 (2006)
5. Groher, M., Zikic, D., Navab, N.: Deformable 2D-3D registration of vascular structures in a one view scenario. IEEE Transactions on Medical Imaging 28(6), 847–860 (2009)
6. Paragios, N., Rousson, M., Ramesh, V.: Non-rigid registration using distance functions. Computer Vision and Image Understanding 89, 142–165 (2003)
7. Lavallee, S., Szeliski, R.: Recovering the position and orientation of free-form objects from image contours using 3d distance maps. IEEE Trans. Pattern Anal. Machine Intell. 17, 378–390 (1995)

8. Lombaert, H., Sun, Y., Grady, L., Xu, C.: A multilevel banded graph cuts method for fast image segmentation. In: Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV), vol. 1, pp. 259–265 (2005)
9. Palágyi, K., Sorantin, E., Balogh, E., Kuba, A., Halmai, C., Erdőhelyi, B., Hausegger, K.: A sequential 3d thinning algorithm and its medical applications. In: Proceedings of the 17th International Conference on Information Processing in Medical Imaging (IPMI), pp. 409–415 (2001)
10. Schneider, M., Sundar, H.: Automatic global vessel segmentation and catheter removal using local geometry information and vector field integration. In: Proceedings of 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro (2010)
11. Broyden, C.G.: The convergence of a class of double-rank minimization algorithms. Journal of the Institute of Mathematics and Its Applications 6, 76–90 (1970), doi:10.1093/imamat/6.1.76
12. Shanno, D.F.: Conditioning of quasi-newton methods for function minimization. Mathematics of Computation 24, 647–656 (1970)

Author Index